The Dissertation Committee for Jagannath Swaminathan
Certifies that this is the approved version of the following dissertation:

SINGLE MOLECULE PEPTIDE SEQUENCING

APPROVED BY
SUPERVISING COMMITTEE:

Edward M Marcotte, Supervisor

Andrew Ellington

Rick Russell

Vishwanath R Iyer

Ilya J Finkelstein

Eric V Anslyn

SINGLE MOLECULE PEPTIDE SEQUENCING

by

Jagannath Swaminathan, B.Tech; M.S

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May, 2015

Dedication

To my family

Acknowledgements

Who and why should I thank anyone? It was I who managed to stay alive and sane; I did all the experiments (well most of it), bought my own coffee, woke up early every day to write this insanely comprehensive, unreadable and seemingly cryptic thesis.

But I couldn't have done this in vacuum. So I begin by thanking the City of Austin for providing such a wonderful place with friendly people, lakes, green belts, festivals, 6th street, Alamo, tacos, shiner beer, haunted houses and a number of events where I have constantly wondered if BYOB is ok! I have done so many fun things, making and checking off bucket list items over these years. But I couldn't have done any of these staying alone.

I needed friends to go out with and do all these activities. So figure 0 is the timeline of my PhD, listing (hopefully) all my friends who have entered my life. If your name is in the figure, you can use it as a coupon that can be redeemed at any time in your life. No questions asked. Just bring it over to me and I shall buy you a beer or coffee. A special offer if you are Mark (the man who always brings on a smile and a disappointing head shake), Alice (forever sure and awesome), Peggy (with her Franklin spirit), Jeremy (illustrator par excellence), HyeJi (always saying a no), Zhihua (forever a conservative), Kevin (must be cooking something in his truck..), Blake (has the party house and is cool), Sundeep, Sucheta, Rachit, Dan, Gabe, Ophelia, Chris, John, John, Jon…damn. I now have to write all your names!

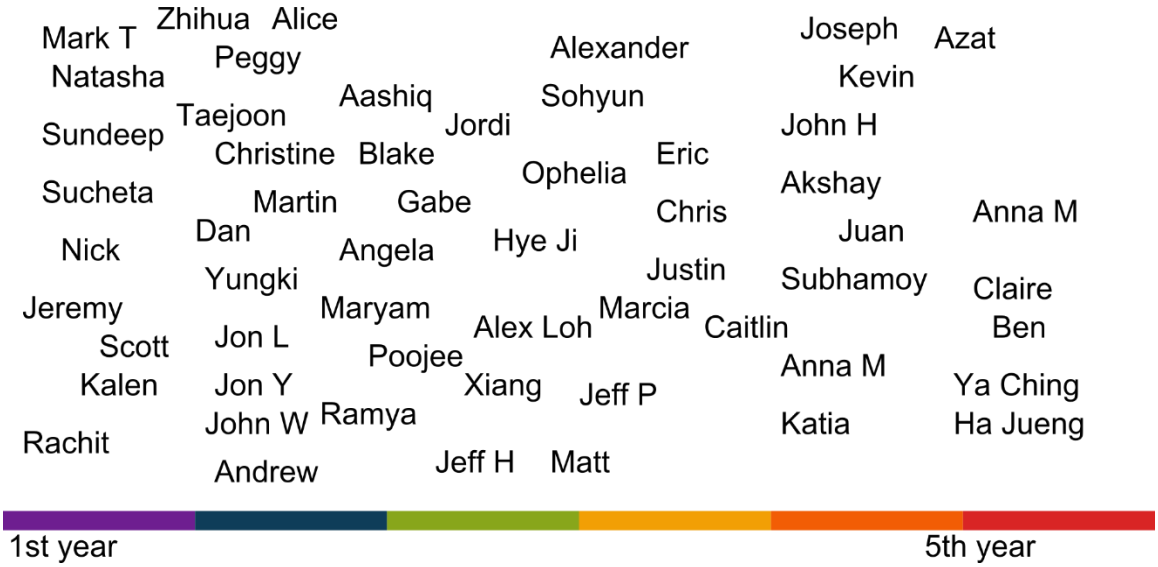

Figure 0: Time line illustrating when I first (approximately) met my friends during the course of my PhD

But then, obviously there were some who actually contributed to make my thesis happen instead of just wanting to see me drunk. I would like to thank Jeff Pruet, Logan,

Amber and especially Erik Hernandez, who I believe, can see through all the evil schemes that the chemical bonds do. Dr. Brian Cannon introduced me to the TIRF microscopy that has lit me up forever. Alexander, the most interesting man in the world, has helped me tremendously with image analysis and seen the entire evolution of the project with me. Joseph Marotta, the engineer on our team, joined us last year and now we all share our cluelessness and advances together. I am also very thankful to William Everett who always helped and indicated that progress was just around the corner. I also like to thank Hugo Celio and Josh Bolinger from the CNM facility who assisted my project in the clean room and with the super-scary XPS machine. My thanks also to Rob Newton, Ashley, Jorge and Barbara for their great administrative help. I thank Steve Fullwood for bringing in microscopes as our Christmas gifts.

My friends during my years in India have shaped me into who I have become and I am grateful for their constant support. Balaguru, Jana, Aslam, Sushanth, Santosh, Chiru, Dushyanth, Aswani Kumar and Pooja. My teachers, Dr K.B.Ramachandran and Dr. Preston Devasia offered a great deal of support.

I also thank my counsellor who helped me get up, when I became disillusioned and disheartened with my life.

But I could never have had this fulfilling PhD, without the ever cheerful, brilliant supervisor, Edward Marcotte; who one cold morning, when there was just me in the lab, came and dropped this fantastic idea onto me! And that's how I did my thesis. I feel I have learnt so much from Edward, like realizing Science is really fun! "If there is no data, it does not exist!" I would also like to thank Andy Ellington for his many deeply inspiring comments, sometimes challenging me to think really big. It was he who had advised and talked me out of doing synthetic biology during the end of my second rotation. I had walked into his office asking for rotation and said that I felt was lost on what I am interested in! In other words, had it not been for Andy Ellington advice that day, I may not have done this thesis. I would also thank my committee members, Eric Anslyn, Rick Russell, Ilya Finkelstein and Vishy Iyer who helped and looked out for me throughout the project.

After thanking so many people, I wonder if it is still my project. I thank my special friends (Chie and Khushboo) and my dear brother (Ramnath) for sticking with me. This thesis is not mine. It's my parents' work (Lalitha and Swaminathan), done through me. This is their PhD.

And finally I would really really really like to thank the first TIRF microscope that our lab purchased for always providing data. Someday, when there will be the rise of the machines and humanity will regret having started this revolution, I will still be grateful to the microscope.

Abstract

SINGLE MOLECULE PEPTIDE SEQUENCING

Jagannath Swaminathan, PhD

The University of Texas at Austin, 2015

Supervisor: Edward Marcotte

The proteome is a highly dynamic and complex set of proteins, specific not only to a particular organism, but to cell types and environmental conditions. Understanding proteome changes as they occur is especially important for molecular diagnostics and developing biomarkers. Currently, the primary technology for proteome wide identification and quantification is shotgun mass spectrometry; while powerful, it lacks high sensitivity and coverage. In this dissertation, I discuss my work in the development of a new technology, termed "fluorosequencing", for sequencing peptides from a complex protein sample at the level of single molecules. The concept is to generate a positional information pattern of an amino acid(s) (such as xKxxK, where K is lysine and x can be any of the other amino acid residues). In order to obtain such a pattern, we proposed a scheme of (i) selectively labeling one or more amino acid(s) in the peptides, (ii) immobilizing millions of these individual fluorescently labeled peptides on a glass surface, (iii) monitoring their changing fluorescent pattern by TIRF microscopy as the (iv) N-terminal amino acid is sequentially cleaved by Edman chemistry and (v) using the

resulting fluorescent signature (fluorosequences) to uniquely identify individual single molecule peptides in the mixture. We began by developing a computational framework to justify the feasibility of the concept. By modeling different sources of anticipated errors, we showed that the errors do not greatly affect the identification of proteins in the human proteome. Secondly, after screening fluorophores for their solvent stability, we used fluorescently labeled synthetic peptides covalently immobilized on beads to experimentally demonstrate the ability of the technique to determine the position of the fluorescently labeled residue in peptides. Finally, we translated the bead optimized chemistry procedures to a single molecule setup. We implemented the fluorosequencing method to sequence synthetic peptide molecules and provided evidence for the technique's utility to discriminate peptides in a peptide mixture with single molecule sensitivity. By establishing the foundational work towards the proof-of-principle for fluorosequencing, we can now scale the method in order to realize the idea of single molecule proteome wide sequencing.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

PROTEOMICS

The proteome is defined as the total set of proteins encoded by the genome and the study of the large scale cellular function determination at the protein levels is a good definition for 'proteomics' [1]. In this thesis, the use of term *proteomics* will be restricted to mean the study of cataloguing the proteins in the protein mixture.

Although with this definition, there would appear an equivalence between the genomic and proteomic information, it is not the case [2]. A number of processes such as regulation of transcription and translation, splicing isoforms, RNA and protein editing [3], half-lives and post-translational modifications such as glycosylation, phosphorylation, truncation etc. are responsible for the cellular presence of thousands of proteins varying in their abundances and/or in their gene-translated sequence. A number of studies have shown weak correlations between the transcriptome (gene expression) and proteome profiles in a number of species [4][1]. These findings indicate that (a) the genomic information cannot be used as a deterministic proxy for cell behavior and (b) alternate methodologies to DNA or RNA sequencing are essential to catalogue the complete proteome datasets. Thus the study of proteomes provides critical and orthogonal information complementing the genomics, transcriptomics, metabolomics and lipidomics of a cell.

Apart from providing fresh insights about biology, proteomics is an important avenue for improving human health [5]. While the origins of many diseases are

[1] There is an ongoing debate in the proteomics community on the degree of correlation between the mRNA levels and protein abundances [4]

Mendelian in nature (see *http://omim.org/*) and can be traced to the mutations in one or a small set of proteins such as in cystic fibrosis [6], Huntington's disease [7] etc., a number of complex diseases involving multi-gene interactions such as cancer [8], cerebrovascular diseases [9], neurodegenerative diseases [10] and infections [11] do not (yet) have a clear identifiable origin of trigger mutations and are best understood and diagnosed by changes in the proteome profile of the cells [12]. In all cases, the symptoms are caused by the disruption of the proteostasis (or proteome balance) of a cell, tissue or system and determining the molecular mechanisms for diagnosis and treatment involves studying changes in their proteomes [13]. Through community led initiatives, a number of human centric proteome datasets are maintained and accessible at the human proteome atlas (*http://www.proteinatlas.org/*), the human proteome project (http://www.thehpp.org/), nextProt (*http://www.nextprot.org/*), the global proteome machine (*http://www.thegpm.org/*, [14] and proteome exchange (http://www.proteomexchange.org/). These information sources are essential for applications in diagnosing and treating diseases. Many of these large scale proteome-wide studies are performed using shot-gun mass spectrometry. Although powerful and modern variants of mass spectrometry have advanced the sensitivity, the depth in proteome coverage is currently low for single runs. Proteins of lower abundances in the sample are not detected [15].

This thesis introduces and describes the foundational principles (theoretical and experimental) for a new technology called "**Fluorosequencing**", in order to perform multiplexed protein sequencing. Analogous to the advancement of genomics with the next-generation DNA sequencing, this technique when scaled, can rapidly identify millions of proteins in complex proteomic mixtures at the level of a single molecule.

CURRENT TECHNOLOGIES IN PROTEOMICS

The technology space for fluorosequencing in proteomics arises due to the lack of existing technologies capable of detecting proteins at high sensitivity and throughput. While early studies in proteomics were pioneered with the use of 2D gel electrophoresis [2], the currently used technologies can be broadly classified as (a) antibody detection based method and (b) mass spectrometry.

In the antibody based detection technology, the specificity of antibody recognition to a protein (antigen) is leveraged. A number of antibody based proteomics platforms capable of identifying about 700 proteins in parallel are commercially available like the planar arrays offered by Panorama technology (Sigma-Aldrich Co, MO, USA) or bead based antibody arrays offered by xMap technology (Luminex, TX, USA). In the global collaborative Human protein atlas program [16], a number of antibodies are developed to specifically target the ~20,000 non-redundant putative proteins. The recent report on the success in profiling ~90% of these proteins across 32 different tissue types with ability to perform localization studies indicates the promise of this technology for large scale proteomic studies [17]. However it is limited to identifying the gene coding proteins, has an inherent difficulty in raising highly specific antibodies against target proteins and their different post-translational modifications and has a difficult problem of displaying proteins in a native conformation for specific antibody binding and detection [18–20].

The use of liquid chromatography coupled tandem mass spectrometry (LC-MS/MS) is the most widely used technology for proteomics today [21]. The use of electrospray ionization technique and other soft ionization methods in mass spectrometry[2] greatly benefited proteomics because liquid protein samples could then be aerosolized

[2] The Nobel Prize in 2002 for chemistry was awarded to Dr. John Bennett Fenn for the development of electrospray ionization and Koichi Tanaka for soft desorption ionization.

and ionized without fragmentation. Along with the development of these nanospray and ionization technologies, the use of in-line high performance liquid chromatography and technologies for isolating and fragmenting individual peptides/proteins were some of the improvements resulting in higher resolution and quantitative protein identification [22]. Further improvements in sample preparation protocols and bioinformatics software has enabled the widespread use of the shot-gun proteomics approach to protein identification in complex protein samples. This method was central to deep profiling of the human proteome [23,24] and in the survey of protein complexes in human cells [25].

Although mass spectrometry has evolved to an atto-mole sensitivity and is capable of obtaining reasonably deep coverage of the proteomes [26], the process of sample preparations and data interpretations are still time consuming and labor intensive. For example, ~2 million MS/MS (tandem mass spectrometry) spectra collected from 72 chromatographic fractions of a human cell line (Hela) could identify only ~45% of the total proteome [27]. This is primarily due to the difficulty in identifying proteins present in low concentration within samples of large dynamic range of protein abundances [28]. In addition, some proteins like receptors, signaling molecules, hormones etc. can trigger signaling cascades at low abundances (in 10s or 100s in numbers per cell). Detecting them is difficult with the current existing mass spectrometers [15]. The fundamental design of detectors used today makes it difficult (if not impossible) to identify these proteins and achieve sensitivity of single molecules. In addition, this mass spectrometry method work by isolating and fragmenting peptide peaks in a serial fashion constraining its throughput to about 10^5-10^6 spectra per experimental run.

With the advent of nanopore technology for DNA sequencing, where DNA polymer threads through a nanoscale pore (made of protein or electrical junction) and generates electrical signatures characteristic of the base pair, attempts to obtain protein or

peptide signatures are underway [29–31]. Since none of these studies have been able to discriminate peptides in a mixture, this methodology will not be discussed further.

The limits in the currently existing technologies, described above, exposes a need for a new method in proteomics that is rapid, scalable with high throughput and is sensitive to detection of even low abundant proteins. Such a technique would ensure the ability to obtain high coverage of the proteome in samples such as cell lysates or blood serum.

CONCEPT OF FLUOROSEQUENCING

The basis for this technique is that the positional information of a small number of amino acid types in a peptide may be sufficiently reflective of the peptide's identity, to allow its identification in a known protein sequence database. **Figure 1** illustrates the proposed scheme for single molecule peptide sequencing. The key idea is to selectively fluorescently label amino acids on immobilized peptides, followed by successive cycles of removing peptides' N-terminal residues (by Edman degradation) and imaging the corresponding decreases of fluorescence intensity for individual peptide molecules. The resulting stair-step pattern of fluorescence decreases will provide the positional information of the select amino acids and often be sufficiently reflective of their sequences to allow unique identification of the peptides by comparison to a reference proteome.

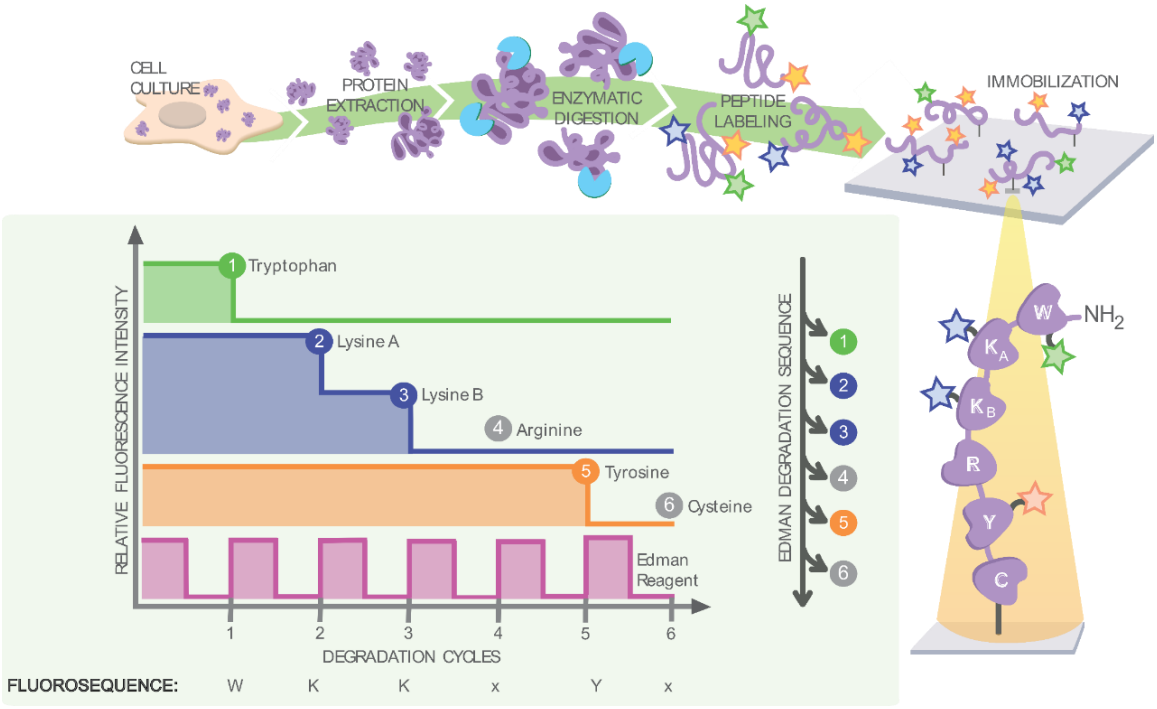

Figure 1.1: Fluorosequencing strategy for single molecule protein sequencing

In brief, proteins are extracted digested into peptides by specific endo-peptidases. All occurrences of the particular amino acids on the peptides are selectively labeled by fluorophores (for example yellow for tyrosine, green for tryptophan and blue for lysine) and surface immobilized for single-molecule imaging (for example by anchoring via cysteine). The peptides are subject to Edman degradation; in each cycle, a fluorescent Edman reagent (pink trace) couples and removes the terminal amino acid The step drop of fluorescent intensity indicates when labeled amino acids are removed, which in combination with the Edman cycle completion signal, gives the resulting fluorosequence (e.g., "WKKxY.."). Matching this partial sequence to a reference protein database identifies the peptide. [Image credit: Angel Syrett]

In more detail, proteins in a complex mixture are first digested into peptides using an endo-peptidase of known cleavage specificity. Select amino acid types (e.g. lysine, tryptophan or tyrosine) are then covalently labeled with spectrally distinguishable fluorophores, each being specific (by reactivity) to the given amino acid side chain. Labeled peptides are immobilized on a glass surface, as for example *via* the formation of

a stable thioether linkage between a maleimide functionalized surface and the thiol group on cysteine residues [32]. The choice of peptidase, labeled amino acids, and anchoring amino acid, all convey information about the identity of a peptide and thus can be optimized for maximum benefit. Using techniques such as Total Internal Reflection Fluorescence (TIRF) microscopy, individual peptide molecules can be imaged on such a surface, and the fluorescence intensity across all fluorophore channels can be determined for each peptide on a molecule-by-molecule basis. By performing Edman chemistry [33] (the classic chemical method used for sequencing amino acids in a peptide by cleaving one amino acid at a time from the N-terminus of a peptide) we can monitor the decreases in fluorescence intensity following each experimental cycle. This informs the relative positions of labeled amino acids in the peptides, and thereby obtain a partial peptide sequence. This scheme can be improved by using a fluorescent Edman reagent coupling and decoupling of which can be observed, enabling the successful completion of each Edman cycle to be monitored for every single peptide, providing an internal error check.

We term the pairing of a single Edman degradation cycle and the subsequent observation for changes in fluorescence as an *experimental cycle*. The observed sequence of luminosity drops in fluorescence across experimental cycles is a *fluorosequence*; the technique itself is thus *fluorosequencing*. For the example shown in **Figure 1.1**, the fluorosequence is "WKKxY". By mapping the partial sequence back to a *reference proteome* of potential proteins, which could be derived from a genome sequence, we would determine if the fluorosequence uniquely identifies a peptide, and ultimately, its source protein.

Commercially available TIRF microscopes can be used to monitor fluorescence changes for millions of individual peptide molecules [34] and are not dissimilar to early variants of next-generation DNA sequencers [35]. By increasing peptide density and

acquiring TIRF images over a large surface area, one could in principle obtain fluorosequences for millions or billions of peptides in parallel. Critically, this approach would be intrinsically quantitative and digital, based on counting repeat peptide observations, in much the same way next generation RNA sequencing is for identifying and quantifying RNA transcripts.

The fluorosequencing technology fundamentally relies on the two established methods – (a) Edman degradation chemistry for the stepwise cleavage of amino acids from the peptide and (b) the use of single-molecule TIRF microscopy [36] for imaging billions of labeled peptides tethered to a glass surface. The relevant background information for these two methods are provided to help the reader understand the underlying principles and limitations of the technology.

METHODOLOGY USED FOR DETECTING THE SINGLE FLUORESCENT PEPTIDE MOLECULES

This dissertation discusses single molecule fluorescence microscopy as the principle technology used for detecting and sequencing peptides. The technique has been developing over the past four decades [34] and has been used in early single molecule DNA sequencing [35] and in studying the dynamics of DNA and protein [37]. A background has been provided to aid the reader in understanding the relevant features and limitations of the detection method used in the fluorescence microscopy.

Principles of fluorescence

Fluorescence was first discovered by George Stokes when he observed that the visibly clear solution of quinine sulfate turned blue when held in the non-visible and beyond the violet portion of the solar spectrum (spectrum separated by a prism)[3] [38].

The Perrin-Jablonski diagram, shown in **figure 1.2**, illustrates the mechanism of fluorescence by tracing the transitions of the excited electrons through the different energy states. Of particular relevance to enhancing the photostability of the fluorophore is the triplet state of the excited fluorophore. This low energy and long lived state (order of micro to milliseconds) occurs when the spin of the excited electron is flipped resulting in delayed fluorescence or phosphorescence. The depletion of electrons from the triplet state can occur by thermal activation or annihilation with other triplet molecules. It is in this triplet state where molecular oxygen (present in aqueous imaging buffers) undergoes an energy transfer with the excited fluorophore resulting in a singlet oxygen species that quenches most fluorophores by oxidation or other covalent modification [39].

Photophysics of organic fluorophores

Most fluorescent compounds are aromatic and/or highly unsaturated organic molecules. In general, the increase in the length of the π-electron system (i.e. the degree of conjugation) shifts the fluorescence spectra to longer wavelengths and increased quantum yield. However, the effect of substitution on the aromatic hydrocarbons is complex and not generalizable for predicting fluorescence behavior. The polarity of the solvent governs the fluorescence behavior due its effect on the dipole moment and relaxation mechanism of the excited fluorophore molecule [40]. Despite the long history of theoretical studies on organic fluorophores, the use of commercial fluorophores

[3] Although he called it dispersive reflexion, in the footnote he termed this phenomenon as "fluorescence" (from combining words - fluorspar, a mineral and opalescence).

(sometimes without the knowledge of their structure) to predict stability or fluorescence behavior under different solvent conditions requires empirical study on a case by case basis.

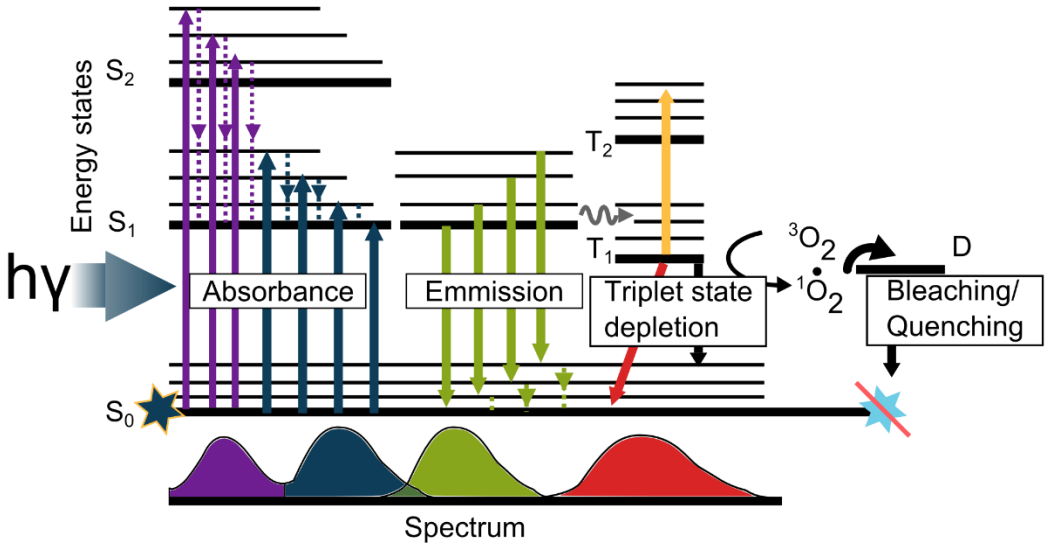

Figure 1.2: Perrin-Jablonski diagram illustrating the mechanism of fluorescence and photobleaching

The fluorophore is excited by the incident laser beam to higher orbital states S1 and S2. A number of internal conversions and subsequent return to the ground state S0 results in fluorescence emission. If the electron spin is reversed in the excited state, the fluorophore enters the triplet state (T1 and T2). Dissolved molecular oxygen in the environment could be converted to the highly reactive singlet state by collision induced energy transfer. This shortens the lifetime of the fluorophore in the triplet state, causes phosphorescence or forces the fluorophore into a long lived dark state (D) where it can be oxidized and bleached or quenched. [The image was adapted from the Wikipedia entry on Jablonski diagram (https://en.wikipedia.org/wiki/Jablonski_diagram).]

Total Internal Reflection Fluorescence (TIRF) microscopy

TIRF microscopy is an optical method for achieving selective illumination of fluorescent molecules in an extremely thin axial section (~50-200 nm) above the glass surface. This enables acquisition of signal from spatially separated single fluorescent

molecule with low background noise. The method is based on the principle that when total internal reflection of the incident laser occurs at the interface of glass-water, a surface electromagnetic field (termed "evanescent wave" and which has the same frequency as the incident light) excites the fluorescent molecules immobilized on the glass surface [41]. The two common setups of the TIRF microscope are the prism mode and the objective mode [42]. In the objective mode (the setup used in this thesis), a specialized objective is used to align the laser beam and achieve TIRF in an inverted microscope, allowing for multi-wavelength laser switching and ease of operating a fluidic device. Single molecule studies using TIRF microscopy have studied reaction kinetics in chemistry [43,44], sequencing of DNA [35], protein dynamics [45,46] and many cellular processes [42].

Photobleaching

The fluorescence emission from fluorophores does not last forever and photo-destruction of fluorophores occur after a period of time. This photo-destruction process, termed photobleaching, is especially problematic for single molecule fluorescence microscopy studies where detecting at least a 100 photons is needed for medium accuracy experiments [46] and fluorescence over time period of minutes is required for studying many biological phenomena [47]. Understanding the phenomena of photobleaching and improving the stability of fluorophores is especially critical for the success of this technology.

The mechanism of photobleaching has been explained by the permanent destruction of fluorescence either through photo-induced chemical destruction or irreversible covalent modification from the excited solvent molecules [46]. An irreversible covalent modification such as oxidation typically occurs by the energy

transfer of the fluorescent molecule in its long lived triplet state with molecular oxygen in the local environment. This can result in either an excited singlet oxygen or the formation of superoxide radical [48,49] that can attack fluorophores [50]. However, since oxygen is also a potent quencher of the triplet state there is a tradeoff between reduced yield due to photo-destruction and increased yield by reduced triplet state population of fluorophores [39].

Although lowering the intensity of excitation light or limiting the exposure is the simplest method to reduce photobleaching [51,52], they reduce the amount of photons being collected making imaging and data analysis difficult. Photobleaching can be reduced by lowering the levels of dissolved oxygen and use of triplet state quenchers. To reduce oxygen in the environment and protect the fluorophores, antioxidants in anti-fade reagents like n-propylgallate in glycerol, ascorbic acid, cysteamine etc. or an enzymatic oxygen scavenger system (OSS) are used for single molecule fluorescence microscopy applications[53,54]. The most common OSS is the glucose, glucose oxidase and catalase mixture that reduces molecular oxygen level to micromolar concentration [55]. This lengthens the photobleaching time for many fluorophores but increases the acidity of the solution. Other systems include protocatechuic acid and protocatechuate-3,4-dioxygenase [53] or pyranose oxidase and catalase [56] can further reduce the dissolved oxygen concentration and does not alter the pH of the solution. Another method of enhancing photostability is to use triplet state quenchers like Trolox [57], β-mercaptoethanol, [58] etc. However many of these chemicals have low water solubility. If the imaging can be performed under non-aqueous conditions, methanol and ethanol would be ideal imaging solvents due to their lower levels of dissolved oxygen and their ability to solubilize increased amounts of triplet state quenchers.

A large number of recent studies have been focused on dramatically improving the photostability of fluorophores by covalently coupling them with protective agents such as cyclo-octatetraene, 4-nitrobenzylalcohol (NBA) or Trolox on cyanine dyes [54,59] or structurally rigidifying groups like azetidene on rhodamine class dyes. These self-healing or ultra-stable dyes, which are typically cell permeable, have enabled observing cellular events such as intra-subunit rotation of bacterial ribosomes or in nuclear staining [54,60].

The design of surfaces and microfluidic devices is another route undertaken recently to improve the photostability of dyes. For example, shielding the rhodamine dye from radicals by introducing them in confined environments like PDMS wells [61], designing PDMS microfluidic device wherein the imaging buffer is deoxygenated by having flanking N_2 channels [62] or by introducing surface modifications [63] have enhanced the photostability of dyes.

METHODOLOGY FOR PEPTIDE SEQUENCING - EDMAN DEGRADATION

The principle method used in the fluorosequencing technology for the stepwise degradation of peptides is a 60 year old chemistry technique called “Edman degradation” [64]. Since the introduction by Pehr Edman , a number of laboratories and facilities in universities house an automated machine that performs Edman chemistry on purified peptide samples (at nano-mole to pico-mole levels) and provides a HPLC chromatogram indicating the sequence of the peptide. An in depth discussion of the basic Edman chemistry, the history of its instrumentation, detection methods and alternative degradation chemistries is presented to aid the readers in the appreciating the nuances of the chemistry used in fluorosequencing.

Chemistry of Edman degradation

In the method of Edman degradation, the amino acid at the N-terminal is labeled with an amine reactive group such as phenylisothiocyanate and cleaved, without disrupting the remaining peptide backbone. The released amino acid is detected and by performing repeated cycles of Edman degradation, the sequence of amino acids is obtained. The principal chemical steps in Edman degradation (shown in **figure 1.3)** are -

1. Coupling: The deprotonated α-amino group at the N-terminal amino acid reacts with phenylisothiocyanate (PITC) under basic conditions. The formed adduct, phenylthiocarbamyl (PTC) peptide, is acid labile when compared to the peptide bond.
2. Cleavage: Treatment with anhydrous acid results in the thiourea attacking the nearest carbonyl group present in the first amino acid causing an intramolecular cyclization. This is followed by the cleavage of the peptide bond adjacent to the carbamyl group while the anhydrous nature of the reaction condition prohibits the cleavage of the remaining peptide bonds. Thus the terminal amino acid derivative is cleaved as a thiazolinone derivative or anilinothiazolinone (ATZ) amino acid) leaving the rest of the peptide intact.
3. Conversion: The cleaved thiazolinone amino acid is selectively extracted by organic solvents and treated with aqueous acids [65] to form a stable phenylhydantoin (PTH) amino acid derivative. The PTH amino acid is detected by different methods, like chromatography, electrophoresis etc. and corresponds to a unique signature of the amino acid variant.

In this way, the N-terminal amino acid is identified and the degradation cycle on the intact peptide is repeated to elucidate the sequence of amino acids on the peptide. Given the wide use of Edman degradation, a variety of conditions for coupling, cleavage,

detection and other modifications in the chemistry or instrumentation has been used (**Table 1.1**).

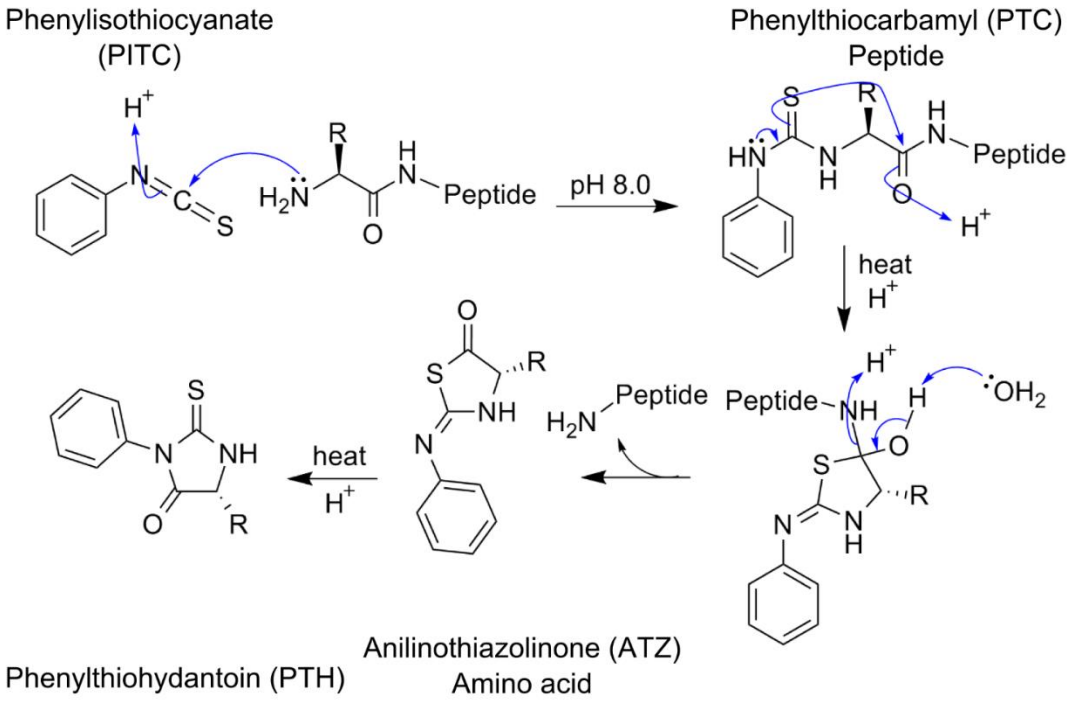

Figure 1.3: Mechanism of Edman degradation

Phenylisothiocyanate (PITC) couples with the amine group on the N-terminus of the peptides under basic conditions to form a phenylthiocarbamyl (PTC) peptide adduct. This undergoes rearrangement under anhydrous acidic condition forming a five membered ring with the first amide bond and cleaves off the peptide with the first amino acid forming an Anilinothiazolinone (ATZ) amino acid. Under aqueous acidic condition, ATZ-amino acid rearranges further to form phenylthiohydantoin (PTH).

In the early years, a number of variants of coupling and cleavage condition were tried by different groups [66], but the use of phenylisothiocyanate became the most popular reagent due to its ideal rates for coupling and cleavage. The mechanism of cleavage was debated in the early years, as to whether water was required for the formation of the PTH amino acid. It was finally established that cleavage occurs under mild anhydrous acidic condition and is a very fast reaction, but conversion of the ATZ-amino acid to the PTH-amino acid required hydrolysis and/or heat with strong acid [66–69]. It was also noted by Edman, that presence of oxygen and oxidative reagents results in oxidative desulfurization reaction of the PTC-peptide, wherein there is a loss of sulfur and a non-cleavable phenylcarbamyl amino acid derivative is formed [69]. Apart from this, it was elucidated that the major loss in efficiency could be due to impure solvents, side reactions or degradation products of PTC-amino acids during the conversion step or loss of peptides from the support [70].

An early attempt in using Edman degradation on a population of peptides (such as those binding to MHC) or synthetic peptides was undertaken by Stevanović and Jung [71], called pool sequencing. They performed Edman degradation on the mixture and from the observed yields of amino acids, back calculated the likely amino acids in the sequence. They used the experimentally determined estimations of the repetitive yield of the peptides in the sequencer and the efficiency in detection as their prior knowledge for identifying the peptide sequences. Although they could not determine the sequences generated from the use of unknown proteases, it was the first attempt in using Edman degradation in multiplexed peptide identification.

Table 1.1: Summary of the major variations in the coupling and cleavage conditions used for Edman degradation.

Coupling condition	Cleavage condition	Identification	Support	Comments	Reference
20% PITC in dioxane incubation for 2-3h at 40 °C	Glacial acetic acid/5.7N HCl mix incubation for 4-16h	PTH-AA or amino acid were analyzed by paper chromatography	Whatman #1 paper strip was used for immobilization and chemistry.	Simplified method of peptides spotted on paper and the chemistry performed on it.	[72]
100:3:1 of Pyridine: triethylamine: PITC incubation at 40 °C for 2 hours with constant mixing.	Mix of 1mL water and 2 mL of HCl saturated with acetic acid incubation for 2 h ours at 40 °C.	Paper chromatography used for separation and UV identification.	Liquid phase chemistry.	One of the early attempt in optimizing Edman degradation for large proteins.	[73]
5% PITC in heptane was added after prior incubation with basic solvent (Quadrol). Incubated at 55 °C for 30 minutes.	Dry HFBA was repeatedly purged.	Gas chromatography of PTH-Amino acid separation.	Spinning cup method used for peptide immobilization and reagent delivery.	Paper describes the automated sequencer by P. Edman based on the spinning cup design.	[74]
20% PITC in acetonitrile incubation for 20 minutes at 45 °C.	TFA incubation for 30 minutes at 45 °C.	Radiolabeled PTH-amino acids separated in silica plates and identified by scintillation counters.	Various functional groups derivatives coated on Polystyrene copolymer beads.	First generation of automated sequencer. Has μmole sensitivity;	[75]

Table 1.2 (contd.): Summary of the major variations in the coupling and cleavage conditions used for Edman degradation.

Coupling condition	Cleavage condition	Identification	Support	Comments	Reference
50% Pyridine and N-ethylmorpholine prior incubation before PITC addition. Incubated for 1hour at 45 °C.	TFA incubation for 20 minutes at 37 °C.	Thin layer polyimide plate used for separation and fluorescence of dansyl amino acid detected for identification.	Liquid phase chemistry;	Only the N-terminal amino acid of the peptide was identified by the dansylation method.	[76]
15µL of PITC with 200 µL of 50% Pyridine incubated for 60-90 minutes at 45 °C.	TFA incubation at 45 °C for 15 minutes.	HPLC or amino acid analysis was performed.	Liquid-phase chemistry;	A cost effective way of Edman chemistry by housing the entire unit under nitrogen	[77]
Ethanol: triethylamine: water: PITC mix (7:1:1:1 v/v) incubation for10 minutes at 50 °C.	TFA purging for 6 minutes at 50 °C.	Analytical HPLC used for separation and identification of PTH-amino acids.	Polybrene membrane used for peptide immobilization.	Manual Edman degradation; Describes other support used as well.	[70]
0.4M Triethylamine in propanol: water (3:2) solvent added prior to PITC. Performed at 55 °C for 20 minutes.	TFA incubation at 40 °C for 15 minutes.	GC and TLC used for separation and identification.	Liquid-phase	The replacement of Butyl chloride with benzene reduced peptide loss. Coupling reagent was maintained under nitrogen.	[65]

Table 1.2 (contd.): Summary of the major variations in the coupling and cleavage conditions used for Edman degradation.

Coupling condition	Cleavage condition	Identification	Support	Comments	Reference
15% PITC in n-heptane incubation for 15 minutes at 42 °C.	TFA for 15 minutes at 42 °C.	HPLC used for separation and identification.	Gas-liquid sequencer introduced.	Polybrene membrane used for sample immobilization. Chlorobutane was used for extracting PTH-AA.	[78]
2.5% PITC (in 50% Pyridine/water) incubation for 20 minutes at 50 °C.	TFA incubation for 10 minutes at 45 °C.	Thin layer paper chromatography used for PTH- amino acid separation.	Liquid phase reaction.	Manual Edman degradation protocol.	[79]
7:1:1:1 of Ethanol: water: triethylamine: PITC at 50 °C for 10 minutes.	TFA incubation at 50 °C for 6 minutes.	Radioactivity of the eluent measured.	Arylamide-sequelon(R) disk used for the solid phase chemistry.	Use of manual Edman sequencing for identifying phosphopeptides.	[80]

Efforts in instrumentation

The simplistic nature of the Edman degradation chemistry enabled optimization of the manual Edman degradation procedure and resulted in the widespread use of PITC as the Edman reagent for coupling and anhydrous trifluoroacetic acid as cleavage reagent [79]. The mechanistic understanding of Edman degradation led to the development of automated sequencers, or sequenators, capable of sequencing long polypeptides up to 60 amino acids in length [74,75]. The sequenators rely on separating the cyclization and conversion reactions in space and time [66], maintaining an oxygen-free environment [77,81] and driving high purity solvents by positive gas pressure.

The first automated sequencer, introduced by Edman and Begg and called the "protein sequenator" used a spinning cup method for effective peptide immobilization and efficient liquid transfers [74]. In the spinning cup design, the peptide and liquids are transferred to the spinning cup where centrifugal forces spread them on the wall. This enables easy liquid phase extractions of the cyclized products. A number of changes to this design resulted in an optimized instrument capable of sequencing a number of proteins or peptides like stomatostatins, interferons etc. at picomolar concentrations [82].

Another automated Edman sequenator was developed around the same time by immobilizing polypeptides on functionalized polystyrene beads and using radiolabeled PITC to aid detection [75]. A number of significant efforts were undertaken to improve the immobilization of peptides on these solid resin. This solid phase Edman degradation step met the criteria set for solid phase chemistry which were that the resins needed to be porous and expandable, provide for high yield in peptide immobilization, lack side reactions, be stable to the solvents used in Edman degradation and be easy to achieve covalent peptide immobilization with the peptide [83].

Schemes for covalent peptide immobilization on amine functionalized polystyrene resins were developed to sequence proteins cleaved at methionine residues [84], lysine residues [85], tryptophan residues [86] and the C-terminus [87]. Other solid substrates used in solid phase Edman degradation were amine coated controlled pore glass [86] and copolymer of dimethylacrylamide which had the ability to expand in aqueous solutions as well [88].

The improvement of instrumentation for Edman degradation has developed from the sequenator designs to micro sequencers and gas phase sequencers. The gas phase sequencer design has been widely used for the last 30-40 years due to its simplified design, ease of operability, sensitivity to about 100 nano-mole peptide concentration and less solvent and reagent use [78]. In gas phase Edman sequencers, the spinning cup design is replaced with a two piece reaction chamber (that houses glass fibers which provides support for immobilization material) and uses volatile solvents. An advantage of gas phase sequencers is that the peptides which are embedded in cationic polymer matrix like polybrene are exposed to volatile solvents like TFA and trimethylamine. This enables retention of the sample in the reaction chambers and allows for high repetitive yields and efficiency. The repetitive yields are greater than 96% and have sensitivity in the pico-mole range [78]. Later efforts involved the use of PVDF membrane to immobilize proteins for sequencing which enabled ease in sample handling [89]. Further attempts in miniaturizing and reducing the amounts of peptides used to atto-mole levels have led to development of a microfluidic platform coupled to an Accelerated MS for Edman degradation [90].

It is evident from a survey of the literature that the early use of Edman degradation was used primarily to sequence a number of proteins and peptides like insulin [91], growth factors, albumins and immunoglobulin [82,91,92]. However, with

the availability of reference genomic sequences aiding tandem mass spectrometry for protein identification, most Edman degradation chemistry is restricted to sequencing novel peptides from organisms lacking genomic information such as snake toxins [93], plant protease inhibitor [94], bacteriocins [95] and antimicrobial or antifungal peptides [95].

Efforts in detection

With the continual evolution of detection technology, the detection scheme for PTH-amino acid has changed from paper or thin layer chromatography [96], gas chromatography [78] to high performance liquid chromatography [97] and mass spectrometry [98]. Currently used automated Edman sequencers are in line with HPLC systems to sequence peptides. With the advent of commercial HPLC systems, a series of alterations of HPLC columns and elution conditions over the 1970-1990s led to the improvement of sensitivity and reproducibility in detecting the PTH-amino acid [97,99–101]. Other efforts in analysis of the N-terminal amino acid involved back conversion of PTH-amino acids (under strong acid or alkaline) to their corresponding amino acids which could then be identified using the widely available and sensitive amino acid analyzers [64,102].

However, due to improved sensitivity offered by mass spectrometry based detection methods, recent efforts were focused on the design and synthesis of kinetically similar isothiocyanate forms, such as dimethylaminopropyl isothiocyanate [103], 4-nitrophenylisothiocyanate, 3-[4′(ethylene-N, N, N-trimethylamino)phenyl]-2-isothiocyanate [104] etc. Formed PTH-amino acid adducts are easily ionizable and can be detected with a sensitivity of femto-moles. An alternative way for using mass spectrometry in conjunction with Edman degradation, is by the use of the “ladder

sequencing" method. In this method, the mass spectrum of the degraded peptide is compared with that of the previous cycle and the difference in the m/z values are used to identify the amino acid [98]. This subtractive method of detection has advantages of being able to detect modifications, has high throughput and does not require high reaction yields [105]. In a fairly recent improvement of a mass spectrometry mode of detection, a radiolabeled [$^{13}C_6$]-phenylisothiocyanate was used on a tryptic digest of BSA and the measurement of the intensity of the isotope labeled amino acid (PTH-amino acid) in the mass spectrum (general technique named as isotope coded affinity tag mass spectrometry) was shown to be applicable for absolute quantitative proteomics [106].

Efforts in development of fluorescent Edman reagents

Fluorescence is potentially more sensitive than UV absorbance as the detection method in HPLC. This has fostered the development of a number of fluorescent Edman reagents over the years to enhance the detection of the thiohydantoin amino acids. The earliest effort was the high UV absorbing naphthylisothiocyanate which however did not have high coupling yields [107]. In the early years a widely touted fluorescent variant was DABITC (4-N,N-dimethylaminoazobenzene 4′-isothiocyanate) that was shown to sequence 23 amino acid at a sensitivity of 2-8 picomoles [108]. Although, this method highlighted the potential of high sensitivity detection of peptides using a fluorescent Edman reagent, the coupling kinetics of the reaction were unfavorable and excess PITC was added along with DABITC to drive the reaction to completion. It has been noted that the major problem of using alternate Edman reagents involve the use of a bulky chromophore that kinetically interferes with the efficiency of coupling and/or cleavage reaction due to the combination of steric and electronic effects. It was later noted that if the chromophore has an electron withdrawing group, it slows down the

cleavage reaction while an electron donating group slows the coupling reaction. It thus turned out that phenylisothiocyanate was the ideal choice as an Edman reagent [66,109]. Another problem in detecting the fluorescent product is that the reagent and the degradation byproducts could also fluoresce and interfere with the chromatographic spectra [110]. Over the years a number of other alternative reagents were used such as FITC (fluorescein isothiocyanate) [111,112], DBD-NCS (7-[(N,N-dimethylamino)sulfonyl]-2,1,3-benzoxadiazol-4-yl isothiocyanate) [113], BAMPITC (4-(N-tert-Butyloxycarbonylaminomethyl)-phenylisothiocyanate) [114], DNTC (4-N,N-dimethylamino-1-naphthyl isothiocyanate) [115] and CIPIC (4-(2-cyanoisoindolyl) phenylisothiocyanate) [116]. An interesting fluorescent Edman reagent is MTBD-NCS (7-methylthio-4-(2,1,3-benzoxadiazolyl) isothiocyanate) which was used in conjunction with Boron trifluoride (BF3) that could differentiate D/L configuration of amino acids by preventing their racemization [110]. While a number of fluorescent Edman reagents have been developed in laboratories and their feasibility in use as an Edman reagent is established, the low efficiency of reaction, commercial unavailability and lack of widespread optimization may explain the lack of use in routine Edman degradation procedures.

Other chemical degradation methods for peptide sequencing

A number of other coupling reagents were developed [66] around the same period of time as Edman degradation. None of the other reagents had as much success as phenylisothiocyanate pioneered by Edman, although a method using thioacetylthioglycolic acid (TATG) as the coupling reagent, seemed to have had moderate success [117]. In this method, the cleaved N-terminal amino acid derivative of

TATG could be converted to free amino acid under mild conditions and analyzed. This method was used to sequence human fibrinogen [118].

An equivalent C-terminal peptide degradation method has been developed and can be used to obtain sequences of blocked N-terminal peptides as well as providing confirmation of sequences obtained by Edman degradation. In this method, the peptide is first activated by acetic anhydride/acetic acid and coupled with thiocyanates to form peptidylthiohydantoins via its carboxylic acid end. The coupling reagent typically used is trimethylsilylisothiocyanate [119]. The peptide bond of the C-terminal amino acid is then cleaved with a base such as acetohydroxamate or triethylamine to form the C-terminal amino acid thiohydantoin and the remaining intact peptide [120]. Unlike Edman degradation which requires a step of conversion, C-terminal sequencing yields a thiohydantoin amino acid which can be readily identified in HPLC. A number of efforts in optimization of the chemistry and automating a solid phase method have consistently been improving the yields of the products, but have not been as widely accepted as the Edman degradation method. The primary problem is the formation of mixed anhydrides resulting in low repetitive yields and the highly selective thiohydantoin formation of only hydrophobic amino acids [121].

CHAPTER DESCRIPTION

This chapter (Chapter 1) has introduced the thesis and the foundational concept for fluorosequencing.

A detailed theoretical study justifying the applicability and scalability of the technique in order to analyze real proteome samples is discussed in Chapter 2. In the chapter, I describe the utility of the technique under ideal conditions and using Monte-

Carlo computer simulations account for the possible errors arising in the method and discuss their potential impact.

In chapter 3, I describe the experimental validation of the fluorosequencing technique in bulk, using synthetic peptides immobilized on beads (solid support). The chapter also describes the choice of ideal fluorophores and their fluorescence characteristics when used for this strategy.

In chapter 4, I provide further experimental proof for the concept and its feasibility in single molecule experiments. This chapter describes the instrumentation, the surface chemistry used, methods for imaging and quantifying peptides and procedures for performing Edman degradation on synthetic peptides at a resolution of a single molecule.

Finally in chapter 5, I discuss future directions of this work.

The appendix describes the characterization and functionalization of an interesting polymer (Cytop) coated on the glass surface.

Chapter 2: Theoretical justification of fluorosequencing

INTRODUCTION

As has been discussed earlier (see **figure 1.1**) and in the publication [122][4], the strategy of fluorosequencing extends the classic Edman sequencing methodology to single-molecule peptide sequencing. By covalently labeling amino acid(s) in peptides with spectrally distinguishable fluorophores, billions of distinct peptides could be sequenced in parallel by Edman chemistry and thereby catalogue the proteins composing the sample and digitally quantifying them by direct counts of the peptides. While it may be clear that having a spectrally distinct fluorophore for every amino acid would directly translate the fluorosequence to its amino acid sequence, it is not experimentally feasible to develop 20 spectrally distinct fluorophores, with each having a high specificity (by reactivity) to a single amino acid side chain. The fluorosequencing technique thus proposes that a limited set of amino acid specific fluorophores (say 1-4) would provide unique positional information of the labeled amino acids in a peptide sequence (such as x-K-x-K, where K is the position of the fluorescently labeled lysine residue and x is an unknown residue). While this kind of pattern is hypothesized to be sufficient in

---

[4] This chapter is adapted from the publication, where Alexander Boulgakov shared the first authorship with me. My role in the publication was in jointly developing the originating ideas along with Dr. Edward Marcotte, performing the theoretical simulation of fluorosequencing and being a part of writing the manuscript. Alexander Boulgakov developed the entire computational framework, helped analyze the data and was part of writing the thesis.

identifying a large number of proteins when mapped to a reference database, it is not evident what the extent of proteome identification would be.

Simulating the fluorosequencing technique computationally determines the extent of proteome identification under the use of different combinations of fluorophores and endopeptidases. With the proposed fluorosequencing method requiring a number of physico-chemical processes such as Edman chemistry, fluorophore labeling etc., different types of errors can be generated and reflected in the observed fluorosequence. Mathematically modeling these errors and providing a computational framework for identifying proteins under different error regimes would provide results on the extent of proteome coverage under these non-ideal conditions. Thus the results from computational simulations would provide the needed theoretical justification for single molecule peptide sequencing in order to multiplex protein identification in proteome samples.

In this chapter, we use the human proteome as an example, to discuss the theoretical considerations for fluorosequencing technique and, by using Monte Carlo computer simulations, explore its feasibility, anticipate the most likely experimental errors and quantify their potential impact on the proteome coverages.

RESULTS AND DISCUSSIONS

Under ideal conditions, even partial amino acid sequences are informative

Computer simulations of variations of this scheme confirm that fluorosequences can be quite information-rich; even relatively simple labeling schemes, employing only 1 to 4 amino acid-specific fluorescent labels, can yield patterns capable of uniquely identifying at least one peptide from most of the known human proteins (**figure 2.1**). For these simulations, we only considered labeling schemes based on known differences in

side-chain reactivity and available amino acid-specific targeting chemistry [123], such as the reactivity of diazonium groups for tyrosine residues [124].

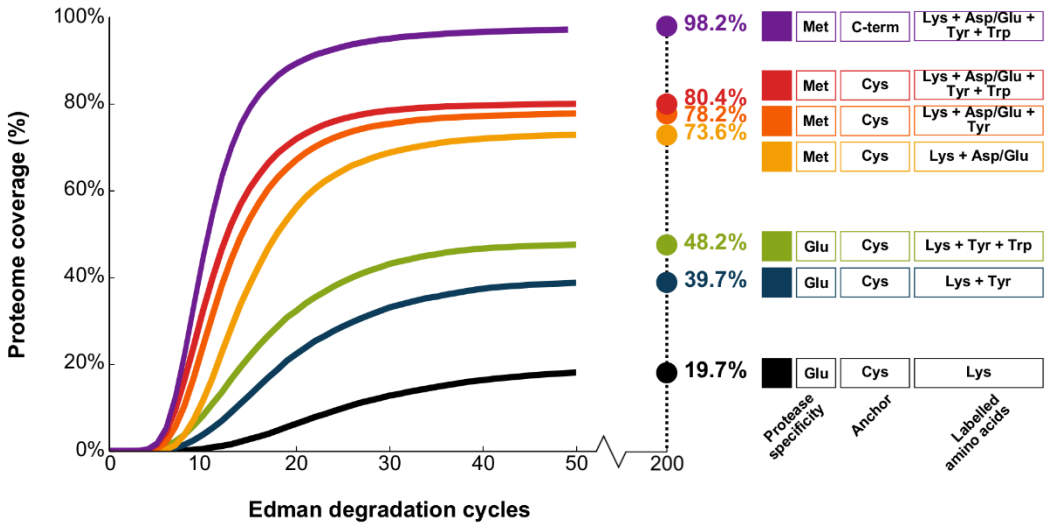

Figure 2.1: Simulations of ideal experimental conditions suggest relatively simple labeling schemes are sufficient to identify most proteins in the human proteome.

Each curve summarizes the fraction of human proteins uniquely identified by at least one peptide as a function of the number of sequential experimental cycles (a paired Edman degradation reaction and TIRF observation). Here, we consider peptides generated by different proteases (*e.g.* Glu represents cleavage C-terminal to glutamic acid residues by GluC, Met represents cleavage after methionine residues by cyanogen bromide) and under different labeling schemes (*e.g.* Lys + Tyr indicates Lys and Tyr selectively labeled with two distinguishable fluorophores. Asp/Glu indicates both residues are labeled with identical fluorophores). Peptides are immobilized as indicated, with Cys representing anchoring by cysteines (thus, only cysteine-containing peptides are sequenced) and C-term representing anchoring by C-terminal amino acids. Increasing the number of distinct label types improves identification up to 80% within only 20 experimental cycles even when only Cys-containing peptides are sequenced; near total proteome coverage is theoretically achievable when cyanogen bromide generated peptides are anchored by their C-termini and labeled by a combination of four different fluorophores. Cycle numbers denote upper bounds, since each fluorosequence is not allowed to proceed past the anchoring residue (cysteine or C-terminus).

Under ideal conditions, many of the labeling schemes (cases where peptides are anchored *via* internal cysteine residues) fail to achieve 100% coverage of the template proteome even after many experimental cycles. The reason is two-fold: (a) Edman reactions does not continue past the cysteine anchor or (b) the proteome contains paralogs and protein families differing at unlabeled amino acids that are hence indistinguishable. When simulations were repeated for the case of anchoring all cyanogen bromide cleaved peptides, not just cysteine-containing ones, by their C-termini, the coverage of the four-label scheme rose from 80% to 98% of the proteome (**figure 2.1, top curve**). Moreover, when simulations were performed for the case of no proteolysis and anchoring each full length protein at its C-terminus, four of the tested multiple-label schemes (including schemes with only 2 label types) achieved over 96% coverage of the proteome within 200 experimental cycles. The remaining proteins were unidentified due to the protein families being indistinguishable using the labeling schemes employed. These simulations thus confirm that single molecule fluorosequencing is intrinsically capable of identifying a majority of proteins in a proteome even when the number of label types is small.

It is also worth considering whether the linear scaling and dynamic range of photon detection by existing cameras might place a limit on the ability to discriminate luminosity drops in fluorescent intensity per peptide. For example, while it might be easy to discriminate a reduction from 5 to 4 fluorophores on a peptide, discriminating a reduction from 25 to 24 fluorophores could be difficult. However, the median count of labelable amino acids per peptide is often small. For example, when considering peptides generated by the protease GluC, this count ranges from approximately 2 (for lysine residues) to 7 (for glutamic acid/aspartic acid residues, which we assume are indistinguishable by reactivity for labeling purposes) (**figure 2.2**). This range is well within the capacity of most modern cameras, since, in practice, TIRF microscopes

equipped with CCD camera variants can count up to at least 13 fluorophores; that is, up to at least 13 copies of a given fluorophore per single molecule can be quantitatively distinguished [125]. Thus, peptides from typical proteomes should not be problematic in this regard.

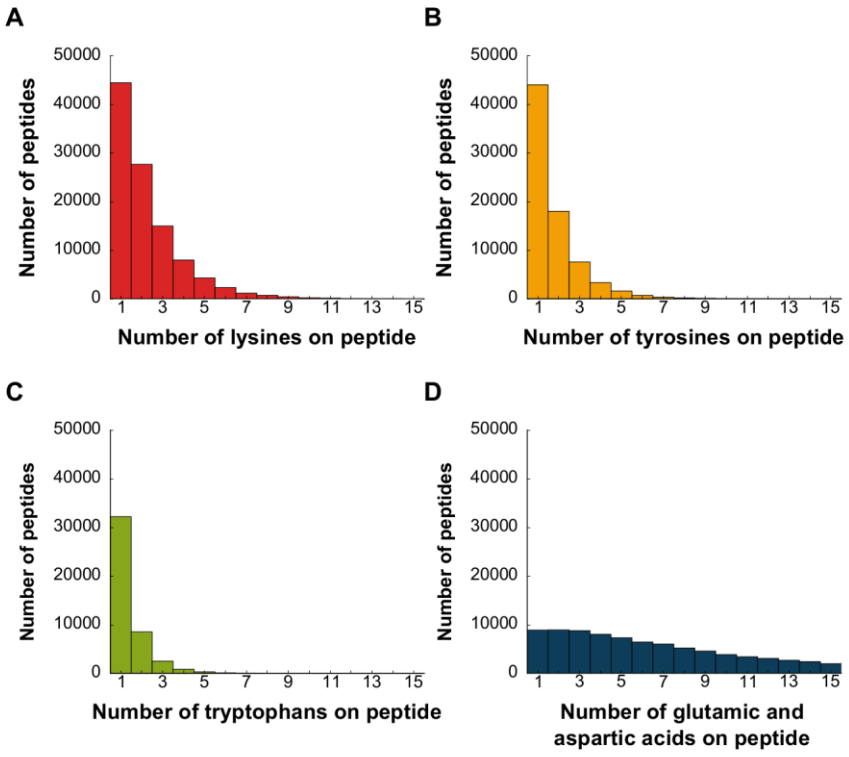

Figure 2.2: Typical proteolytic peptides have counts of labelable amino acids sufficiently low to sequence.

Frequency histograms of amino acids in in-silico proteolytic peptides for lysine (A), tyrosine (B), tryptophan (C), and glutamic acid/ aspartic acid (D) indicate low median values. Peptide sequences in A-C were generated in silico from the human proteome by GluC digestion, and those in D by cyanogen bromide digestion. Low counts of labelable amino acids per peptide are expected to increase the ability to discriminate removal of one fluorophore amongst many on a peptide.

Anticipating the inevitable failures of dyes and Edman chemistry

Being a physico-chemical process, we can forsee some of the most likely potential sources of error for an experimental implementation of the scheme. With errors, an observed fluorosequence would not reflect the true sequence of fluorescently labeled amino acids. Three of the most probable error sources are as follows:

(a) **Failure of fluorophore attachment or emission causing apparent substitutions**: Steric constraints of peptides or reaction kinetics of fluorophore labeling chemistry might result in specific amino acid(s) not being covalently labeled. This scenario is equivalent to correctly coupled but non-emitting fluorophores, such as those observed in defective fluorophores [126]. In both circumstances, the position of a labelable amino acid would be misinterpreted as containing a non-labelable amino acid, *e.g.* the peptide "GK*EGK*" (where K* represents a labeled lysine) would mistakenly yield a fluorosequence "xxxxK" instead of "xKxxK", for a dye failure at second position at lysine.

(b) **Photobleaching of labeled fluorophores causing apparent coupled double substitutions ("residue swaps")**: Another complication in the analysis could arise from the permanent photochemical destruction of dyes. In this scenario, a labeled residue at one position is misinterpreted as an unlabeled residue because the label is lost by photobleaching, while another residue upstream in the peptide (typically unlabeled) is misinterpreted as being labeled because the photobleaching fluorophore loss coincides with that particular experimental cycle. This would shift the apparent position of the label upstream in the fluorosequence. For example, peptide GK*EGK* might be observed as xKKxx when the dye on the lysine at the fifth position photobleaches during the third imaging cycle. This situation reduces the ability to (i) reliably count the number of fluorophores lost during an experimental cycle, (ii) distinguish whether a change in

luminosity results from fluorophore loss due to a genuine Edman degradation step or photobleaching, and (iii) identify which downstream fluorophore was extinguished if the loss is indeed due to photobleaching. Although fluorophore half-lives can be extended by use of oxygen scavenging systems [53], synthesis of stable dyes [54,59,60] or even surface modification [62,63], photobleaching is still a stochastic process and accounting for loss of fluorophores erroneously coincident with upstream Edman degradations would be critical to identification. Currently, there are many photo-stable dyes on the market. A recent study on the effects on dyes by oxygen radicals found that the half-life of Atto647 was roughly 3 minutes (corresponding to 180 experimental cycles at 1 second/cycle exposure) [49], while Atto655 showed a mean photobleaching lifetime of 8-20 minutes [127], corresponding to many hundreds of experimental cycles.

(c) **Inefficiency of Edman degradation chemistry causing apparent insertions**: Optimization efforts of Edman degradation reactions over the past sixty years have resulted in efficiencies of >95% [128]. Nonetheless, failed cycles are expected at some non-zero rate and would yield an observation corresponding to no fluorescence change, even if there was a labeled amino acid in position to be removed. This corresponds to an apparent insertion of a non-labeled amino acid into the fluorosequence. The use of a fluorescing Edman reagent (e.g., DABITC or FITC [129]) would enable direct monitoring of every coupling and decoupling step of the chemistry, providing an internal error check for successful completion of the Edman cycle as in **figure 1.1**. Non-fluorescent Edman reagents such as phenylisothiocyanate are much more commonly used, and we therefore investigated the effect of this parameter.

A framework for modeling single-molecule sequencing under non-ideal conditions

To analyze how the peptide sequencing efficiency is affected by the above three types of errors and to map fluorosequences to source proteins, we developed a modeling framework to simulate the process. Unlike the ideal case where fluorosequences are faithful to their source peptides, and hence mapping to the reference proteome is trivial, accounting for errors such as the three previously highlighted complicates mapping. For example, the fluorosequence "xKxxK" cannot be uniquely attributed to the "GK*EGK*" peptide, since Edman failure at the first position of peptide "K*EGK*" or a fluorophore failure on the first lysine of "K*K*EGK*" could also yield the same pattern. While errors arising from the inefficiency of Edman chemistry and fluorophore failure are tractable by analytical solutions, the non-Markovian nature of photobleaching events forces us to employ a Monte Carlo approach.

We therefore developed a Monte Carlo procedure to simulate thousands of copies of each of the 20,252 proteins in the human proteome being subjected *in silico* to fluorosequencing in order to obtain a random sample of the fluorosequences produced for a specified set of error rates. **Figure 2.3** details the simulation steps; the Methods provide more complete descriptions of the error models and pseudo-code for the overall procedure.

Each sample observation generated by the Monte Carlo simulation is a sequence of luminosity drops yielded by one individual peptide subjected to *in silico* Edman cycles. We conservatively assume that we cannot observe or infer the absolute number of fluorophores labeling a peptide, but that we can monitor and statistically discriminate whether, after each attempted Edman cycle, there has been a decrease in luminosity in each fluorescent channel, consistent with signals previously shown to be discernible for single molecules [125].

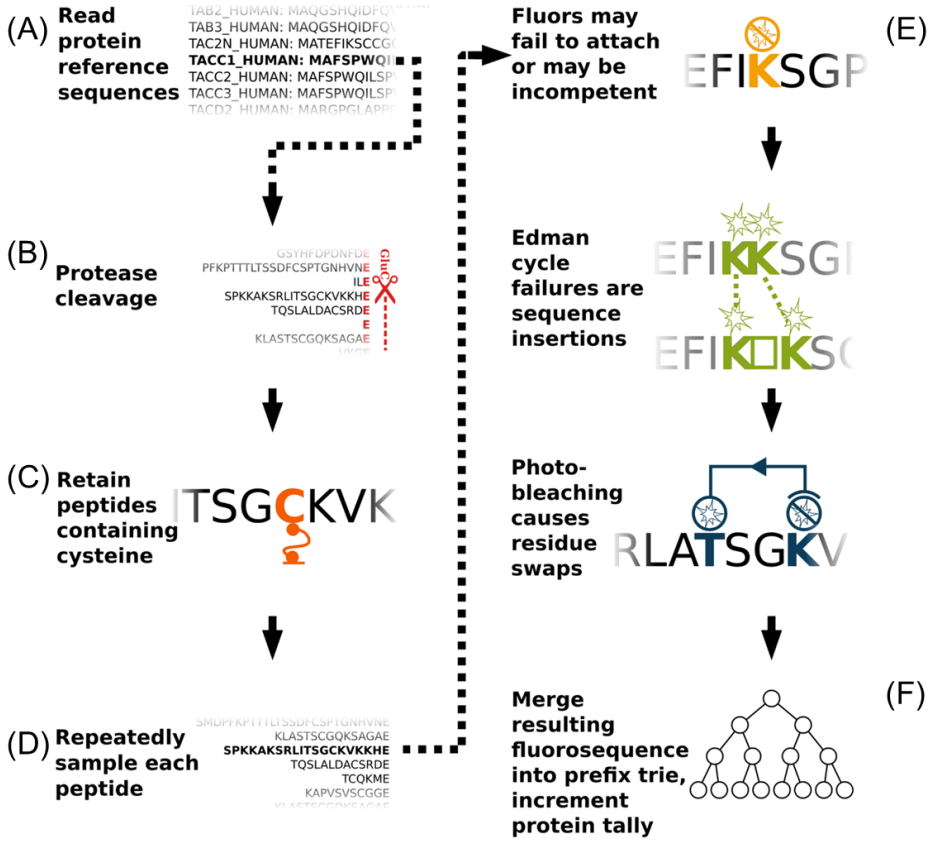

Figure 2.3: Overview of a Monte Carlo simulation of fluorosequencing with errors.

In detail, (A) protein sequences are read as amino-acid character strings from the UniProt database. For each protein sequence, the subsequent steps are repeated: (B) proteolysis was simulated and (C) peptides lacking the residue for surface attachment (e.g. cysteine) were discarded. (D) All remaining peptides were encoded as fluorosequences and subsequent steps were repeated in accordance to the desired sampling depth: (E) The fluorosequences were altered *via* random functions modeling experimental errors - (1) labels were removed modeling failed fluorophores or failed fluorophore attachment, (2) positions of the remaining labels were randomly dilated modeling Edman reaction failures, and (3) fluorophores were shifted upstream from their positions, modeling photobleaching. (F) Each resulting fluorosequence was sorted based on its position and label type and merged into a prefix trie to tally the frequencies of observing each fluorosequence from a given source protein.

For the purpose of simulation, we make the simplifying assumptions that different fluorophores have fully distinguishable signals, do not exhibit dye-dye interactions or Förster resonance energy transfer, nor exhibit channel bleed-over.

The fluorosequences (observed reads) from the simulations are next collated into a prefix trie [130], as illustrated for a simple toy example (**figure 2.4)**. Each fluorosequence is linked in the trie to its source protein(s) and associated count(s) of observations over the course of the simulation, thereby empirically estimating the fluorosequence's source protein probability distribution.

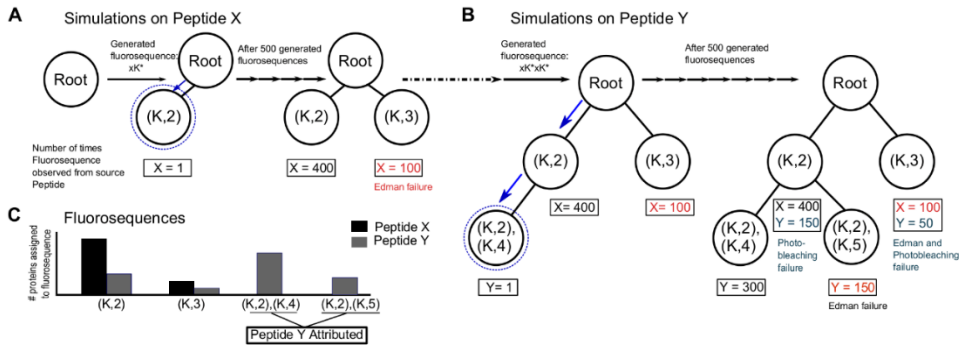

Figure 2.4: A simple example of the trie structure for storing and attributing fluorosequences to peptides or proteins.

Consider a toy peptide mixture with peptide X (sequence GK*EGC, where K* represents fluorescently-labeled lysine; the sequence can be simplified to (K,2)) and peptide Y (GK*GK*EC; represented as (K,2),(K,4)). Panels (A) and (B) summarize populating the trie with fluorosequences from 500 copies each of Peptide X and Y, respectively. For example, peptide X might generate fluorosequence xK*, incorporated into the trie as a new node (K,2). (B) Simulations on Peptide Y add additional nodes to the trie like (K,2),(K,4) after traversing node (K,2). Additional fluorosequences similarly generated are incorporated into the trie along with a tally of the number of observations of each fluorosequence. Following the Monte Carlo simulation, the frequency of each source protein can be calculated for each trie node and thresholds applied to identify and count those source proteins most confidently identified. Here, fluorosequences ((K,2),(K,5)) and ((K,2),(K,4)) confidently identify peptide Y, while Peptide X is not attributed.

Figure 2.5 illustrates two extreme cases of protein probability distributions for a given fluorosequence. Importantly, modeling the frequency of source proteins for fluorosequences is equivalent to obtaining (within sample error) the posterior probability mass functions – *i.e.* the set of probabilities $P[p_j|f_i]$ such that given an observed fluorosequence f_i, the probability that protein p_j is its source (henceforth called the *attribution probability mass function (p.m.f.)*). Notably, by sidestepping problems associated with developing algorithms for inverting fluorosequences to their source peptides, and the peptides' own derivation from source proteins, we make the strategy amenable for incorporating additional experimental parameters, including fluorophore spectral channel bleed-over or protease inefficiencies. Thus, the attribution p.m.f.'s provide a natural framework both for modeling errors and for directly mapping actual experimentally observed fluorosequences to proteins in the proteome. Based on the properties of this distribution, a fluorosequence can be associated with the protein most likely to yield it, for example applying a confidence threshold (see Methods).

In future applications using attribution p.m.f.'s to interpret fluorosequencing data from real samples, one might also wish to model realistic numbers of copies per protein processed through the simulation pipeline, since the Monte-Carlo based deconvolution of fluorosequences to source proteins will be affected by protein abundance dynamic range as well as simulation depth. For example, high simulation depth would not only reduce the sampling errors, but also accurately attribute low abundance proteins from confounding high abundance proteins that generate the same fluorosequence by a low probability event. Although we did not explore this aspect, simulating protein copies based on their prior known abundances [27] might significantly reduce Monte-Carlo simulation computational resources. The version of the simulation we performed here makes no such assumptions about protein abundance, and thus corresponds to a Bayesian

flat prior expectation on protein abundance, applicable to any sample. Under real experimental conditions, we may want to at least simulate an order of magnitude of the abundance for each protein in the proteome to reduce the inaccuracy in identification and quantitation.

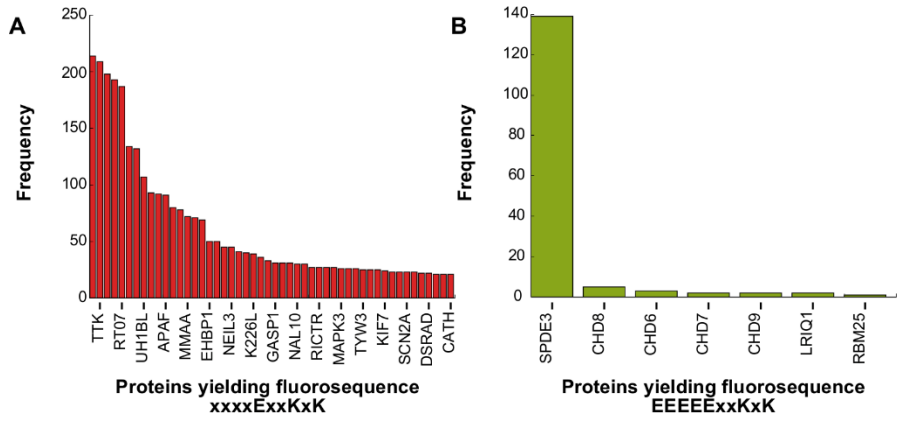

Figure 2.5: Monte Carlo sampling reveals the confidence with which fluorosequences can be attributed to specific source proteins.

(A) and (B) represent two example fluorosequences, illustrating opposite extremes in terms of the number of proteins capable of yielding each sequence. In (A), the frequencies with which rival source proteins yield fluorosequence "xxxxExxKxK" in the Monte Carlo simulations indicates low confidence in attributing that fluorosequence to any one protein. In (B), a single protein is by far the most likely source of fluorosequence "EEEEExxKxK". (X-axes represent incomplete lists of proteins, ranked ordered by the frequency they are observed to generate the given fluorosequence in the simulations.)

More amino acid colors compensate for photobleaching and poor Edman efficiency

Using the Monte Carlo scheme, we simulated sequencing the human proteome to a simulation depth of 10,000 copies per protein, performing a parametric sweep of 216 experimental parameter combinations (corresponding to six values for each of the three error parameters). **Figure 2.6** illustrates the effects for three alternate labeling schemes of varying Edman efficiency and fluorophore half-life on the percentage of proteins

identified after 30 Edman cycles, given fluorophore failure rates ranging between 0 and 25%. As in **figure 2.1**, diversifying the labels offers the greatest improvement in proteome coverage, even with relatively poor process efficiencies.

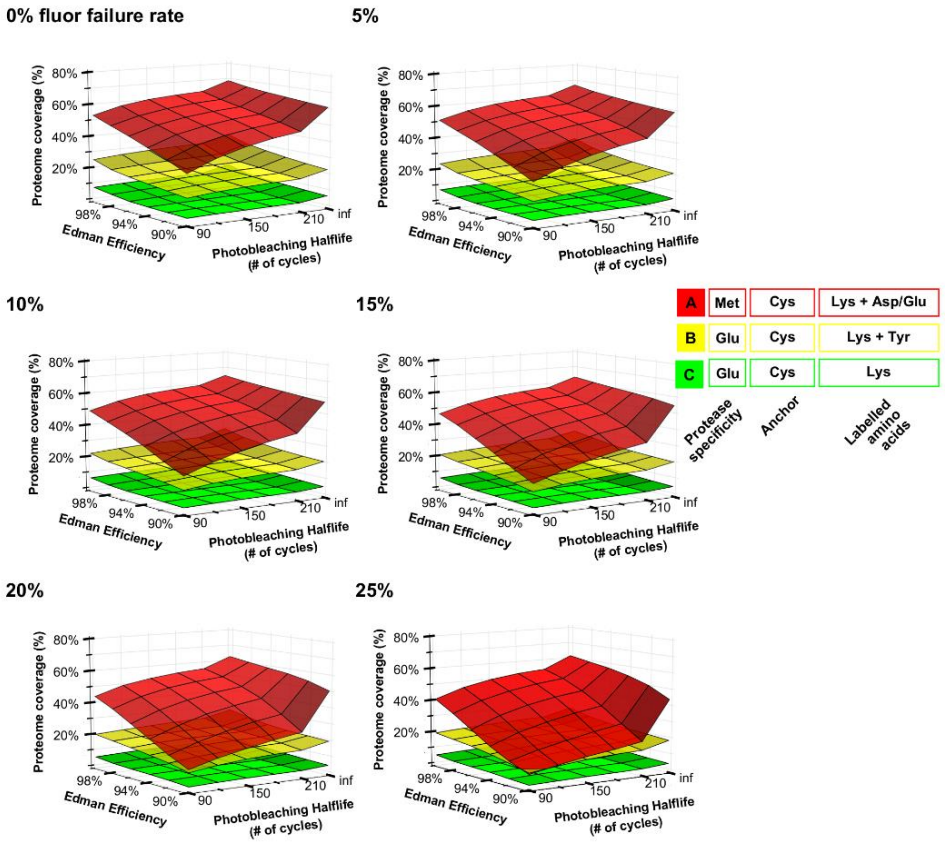

Figure 2.6: Surface plots illustrate the consequences of differing rates of Edman efficiency, photobleaching, and fluorophore failure rates.

Each panel summarizes the consequences of varying rates of photobleaching and Edman failures for a different fixed fluorophore failure rate, ranging from 0% to 25%, as calculated after simulating 30 experimental cycles on the complete human proteome at a simulation depth of 10,000 copies per protein. Photobleaching shows the strongest negative impact on proteome coverage when compared to other errors; increasing the number of distinguishable labels strongly increases proteome coverage. Labeling and immobilization schemes are denoted as in Figure 1.1.

The number of proteins identified is reasonably robust to changes in fluorophore failure rates. For example, a 25% increase in failure rate causes only a 0.8%-6.4% reduction (range includes all parameter combinations) in proteome coverage for schemes B and C (see **figure 2.6** for scheme descriptions). However for scheme A, a 25% increase in fluorophore failure rate causes a 19% reduction in proteome coverage under moderate estimates of photobleaching and Edman efficiency. Scheme A is less robust *vis-à-vis* all simulated errors because the boost in the positional information stemming from abundant aspartates and glutamates is rapidly undermined by experimental errors, as there are higher chances for fluorescently indistinguishable peptides to confound the fluorosequence.

Notably, the photobleaching half-life has the greatest effect of any of the tested parameters on protein identification, causing up to 50% loss in proteome coverage (under scheme A). The steepest decrease in the number of proteins identified occurs when photobleaching is considered (comparing half-lives of infinity to 210 cycles) and tapers with lower half-life. Although photobleaching shows the strongest impact of any of the errors considered, it is worth noting that the half-lives of commercially-available fluorophores are sufficiently longer than those simulated. Hence, we anticipate that this error source will not prevent a real implementation of fluorosequencing. For comparison, literature evidence suggests that common failure rates of fluorophores may be about 15-20% [126,131], Edman degradation proceeds with about 94% efficiency [128], and the mean photobleaching lifetime of a typical Atto680 dye is about 30 minutes [127], corresponding to 1800 Edman cycles, assuming 1 sec exposure per Edman cycle. Thus, we expect error rates to be sufficiently low for effective fluorosequencing.

Determining the positional information of amino acids as a general principle for next-generation protein sequencing

Fluorosequencing relies on the positional information of specific subsets of amino acids within peptide sequences. The scheme can be generalized as a framework fulfilling two conditions – (a) *an observable event* 'e', which occurs by detection of a known single amino acid or a class of amino acids, and (b) *a sequential analytical process*, which increments or decrements the sequence in a known direction and by constrained number of amino acids. While we have suggested using detection of fluorescently labeled amino acids as the event, other modalities might be considered, such as detecting voltage changes or reactivity of monitored amino acids. Besides Edman degradation, other valid sequential processes could include sequential treatment with known sequence specific peptidases or directional protein translocation through a nanopore channel [29] at a defined translocation rate. The monitoring of sequenced detection events gives information-rich patterns (such as "x-e-e-x…" where 'x' is one or more non-identifiable amino acids) capable of being mapped back to a reference proteome. The nature of this information lies between the extremes of information content, wherein either every amino acid corresponds to a distinct event or there is no observable event associated with the process (as, for example, a peptide translocating through a channel but not generating a detectable signal). In principle, many event-process strategies might be suitable for peptide sequencing and interpretation using a scheme similar to the one we present.

CONCLUSIONS

We propose a strategy for the parallel identification of proteins in a complex mixture based on the positional information of amino acids in peptides. The integration of a 60-year-old, highly optimized Edman chemistry [64] with recent advances in single-

molecule microscopy [37] and stable synthetic fluorophore chemistry [60] makes this strategy particularly amenable for experimental execution in the near future. Modeling of experimental errors suggests this strategy can be reasonably expected to identify a high percentage of the proteome, comparable to mass spectrometry, and potentially brings the advantages of single molecule sensitivity and—if next-generation single molecule sequencing is a reasonable proxy—throughputs of hundreds of millions or billions of molecules sequenced per run. Monte-Carlo simulations provide a framework to accommodate the inevitable experimental errors and probabilistically identify proteins from the observed fluorescent patterns. Successful experimental execution of the proposed strategy will not only lead to progress in proteomics, but enable progress in engineering and chemistry to enable the technology.

MATERIALS AND METHODS

Datasets

The UniProtKB/Swiss-Prot complete *H. sapiens* proteome (manually reviewed) was downloaded on 29th May 2013 and used for all simulations, comprising 20,252 protein sequences and ignoring alternatively spliced isoforms.

Monte Carlo simulations

Simulations were programmed in Python using Merseinne Twister as the source of randomness, and implemented in parallel using the Texas Advanced Computing Center. For the purposes of simulation, the proteome can be considered dictionary pairs of protein identifiers and amino acid sequences. We began the simulations with 10,000 copies of each protein sequence. The first two steps in the simulation split each amino acid sequence string at residue(s) corresponding to the protease specificity (*e.g.* E for the

GluC protease) and then discard substrings that lack the anchor residue (*e.g.* substrings not containing C). Alternating Edman degradation steps and TIRF observations on the resulting peptides provide temporal ordering for luminosity drops, resulting in an observed fluorosequence for each peptide. In the simulation, fluorosequences were initialized from amino acid substrings' correct fluorophore positions, and experimental errors were then introduced sequentially, modifying the fluorosequences in accordance with each type of error's appropriate probability distribution.

The three experimental sources of error sources were modeled in the Monte Carlo simulation as follows:

(a) Inefficient dye labeling - The probability of an amino acid not being labeled with its intended label or being labeled with a nonfunctional dye (*i.e.* a dye that attaches but is incapable of fluorescence) is modeled as a Bernoulli variable. For each label prepared for the experimental procedure, there is a probability u that the fluor will never be observed.

(b) Edman degradation is represented as an attempt to remove one amino acid residue per cycle. These attempts are modeled as a Bernoulli process, since every experimental cycle is independent of the preceding cycle. The probability of the N-terminus amino acid being successfully cleaved off is assigned a parameter p and the corresponding failure follows as $q = 1 - p$. Failure of Edman chemistry delays the removal of a downstream labeled amino acid by one experimental cycle, and thus dilates the inter-label intervals in the fluorosequence. Using this model, the probability that an inter-label interval d requires $d + e$ experimental cycles before the subsequent label is removed is $\binom{d-1+e}{d-1}p^d q^e$. A random number is drawn from this distribution to

indicate the dilation for each interval. We assume that Edman chemistry stops at the first cysteine from the N-terminus.

(c) Photobleaching is the irreversible photo-induced destruction of a fluorophore. The photobleaching process can be best described as a stochastic phenomenon and modeled by an exponential decay function [48]. Every fluorophore has a defined half-life based on solvent conditions and laser operating conditions [51]. The periodic laser excitation has an additive effect on the fluorophore's half-life: exciting a fluorophore once for thirty seconds and, after an arbitrary delay, again for a further thirty seconds will photobleach the fluorophore with the same probability as a continuous excitation for one minute. We assume a constant period of laser exposure per experimental cycle. To model whether labeled amino acids have been cleaved, the probability of a fluorophore still on the peptide surviving k experimental cycles can be modeled as an exponential decay e^{-bk}, where b is an experimentally-determined characteristic constant of the fluor being used, k is the number of experimental cycles performed, and e is Euler's constant. We shift labels to earlier experimental cycles based on random numbers drawn from this exponential decay.

For a given simulation, all simulated fluorosequences were collated into a prefix trie whose keys were the sequences of luminosity drops and associated values represented the counts of source proteins yielding those fluorosequences. One trie was generated for each given choice of error rates, protease and labels, based upon simulating 30 Edman cycles of fluorosequencing 10,000 copies of each protein in the human proteome. For each fluorosequence in the resulting trie, its source proteins were counted, allowing proteome coverage to be calculated.

The simulation can be summarized as a pseudo-code:

INITIALIZE **result trie** as an empty prefix trie.

FOR **protein** IN **proteome**:
    peptides ← Proteolyze **protein** at the carboxyl side of a given amino acid corresponding to the protease used.
    FOR **peptide** IN **peptides**:
        Discard **peptide** if it does not contain at least one occurrence of the amino acid for anchoring to the surface.
    FOR **peptide** IN **peptides** REPEAT **10000** TIMES:
        Attach labels to amino acids with a given probability. Labeling probability is uniform and mutually independent for all amino acids.
        Adjust positions of labeled amino acids to reflect possible Edman failures. All Edman reactions for each individual peptide have a uniform probability of success specified by a given parameter, and are mutually independent. We assume the Edman reaction cannot proceed past the first amino acid anchored to the surface.
        Adjust positions of labeled amino acids to reflect potential photobleaching. All fluors' survival functions are mutually independent exponential decays characterized by a given photobleaching constant.
        Collate final sequence of **tuples (fluorosequence)** for this **peptide** into the **result trie**.

TRAVERSE THE TRIE. For each node, find the most frequent source protein yielding that fluorosequence. For the purposes of data visualization, if the most frequent protein yielded the fluorosequence at least ten times, and all other source proteins for that fluorosequence combined are responsible for less than 10% of all observations, then that fluorosequence is considered to be uniquely attributed to the protein.

RETURN the set of proteins that have at least one uniquely attributed fluorosequence.

A parameter sweep was performed for the three labeling schemes as in **Figure 2.6** at a simulation depth of 10^4 copies per protein, sweeping 216 experimental parameter

combinations (testing six values for each of the three error parameters described) spanning fluorophore failure rates of 0%, to 25%, photobleaching half-lives from 90 minutes to infinity (*i.e.,* no photobleaching), and Edman degradation efficiencies from 90% to 100%.

Attributing fluorosequences to peptides and proteins

For more efficient use of computer memory, trie structures were calculated separately for multiple subsets of the proteome and the resulting tries merged before analysis by traversing all fluorosequences in each trie and adding each fluorosequence along with its protein counts into a master trie for that simulation. Then, the counts of each fluorosequence and affiliated peptides were analyzed to calculate a frequency distribution of the number of times peptides from a given source protein generated a given fluorosequence. For the purposes of summarizing the data, two criteria were applied to this distribution to attribute a fluorosequence uniquely to the protein: (a) its primary source protein yielded the fluorosequence at least 10 times out of a 10^4 simulation depth, and (b) the summation of frequency from all other source proteins were responsible for less than 10% of that fluorosequence's occurrences. While the former criterion addresses sample error, the latter addresses confounding from other proteins.

Chapter 3: Fluorosequencing of peptides on beads

INTRODUCTION

The fluorosequencing strategy for single molecule peptide sequencing (described in chapter 2) proposes the method of identifying peptides based on the position of its fluorescently labeled amino acid. This can be achieved by detecting the decrease in the peptide's fluorescence intensity (coinciding with the position of labeled amino acid) through the amino acid cleavage steps of Edman degradation chemistry. In this chapter, which contains experimental designs from collaborative efforts[5], enablement of this technique is described. This technique requires the establishment and optimization of important underlying chemistry procedures. Two of these procedures include (a) immobilization of fluorescent peptides on solid supports and (b) performing Edman chemistry to cleave one amino acid at a time from its N-terminus. While Edman degradation on immobilized peptides has been developed extensively on solid support [75,87,117], the requirement of using fluorescently labeled peptides and detecting their fluorescence changes provide a unique set of challenges.

Despite the long history of the studies on synthesis of fluorophores, it is not evident whether subjecting the fluorophores (especially some of the commercially available fluorophores such as Atto647N, Alexa680 etc. with their superior quantum yields and publically unavailable structures) to Edman conditions will alter their inherent photo-physical properties. Although there is precedence for the use of some fluorophores such as fluorescein isothiocyanate (FITC), 4-N,N-dimethylaminoazobenzene 4'-

[5] Erik Hernandez (from Dr. Eric Anslyn's lab) synthesized all the fluorescently labeled peptides used in this study. Erik Hernandez and Dr. Jeff Pruet helped in providing solution for an important problem in the chemistry.

isothiocyanate (DABITC) etc. [129] as Edman reagents, there is no generalizable structural patterns that can be applied to shortlist fluorophores which could then be used experimentally. Empirically testing the fluorophores for their stability is an experimentally feasible route to narrow down the list of ideal fluorophores for the fluorosequencing technique. While Edman degradation has been optimized to work with all the different amino acid side chains and even glycosylated side chains [132] with relatively high efficiency of >90% [128], it remains to be tested if the presence of bulky and charged fluorophore on the amino acid side chains hinder the reaction. Performing Edman degradation on synthetic peptides with known position of the fluorophores can be used to determine the efficiency of cleavage of the fluorescently labeled amino acid.

We reasoned that testing the Edman degradation process on the bulk fluorescently labeled peptide would indicate the feasibility of the chemistry steps of fluorosequencing. Given the diversity of functional groups on commercially available beads, we decided to use Tentagel beads as the platform for immobilizing fluorophores or peptides, optimizing the chemistry and by image acquisition and processing, quantitate the fluorescent peptide density (see **figure 3.1**). Among the number of other commercially available beads such as controlled pore glass, magnetic beads, polystyrene beads etc., Tentagel beads have a set of advantages for this study due to their compressibility (suitable for imaging by sandwiching them between glass slides), high peripheral density of functional groups (enables quantitation of bound peptides and discriminating the non-specifically attached peptides) [133] and availability as micron sized beads (facilitating imaging and ability to be retained in many fritted syringes).

In this report, we primarily use amine functionalized Tentagel beads to shortlist the fluorophore choices that are ideal for performing fluorosequencing, establish the scheme for immobilizing peptides to the bead via their carboxyl termini and by

optimizing the Edman degradation procedure, provide evidence for discriminating multiple peptides based on the position of their fluorescently labeled lysine residues.

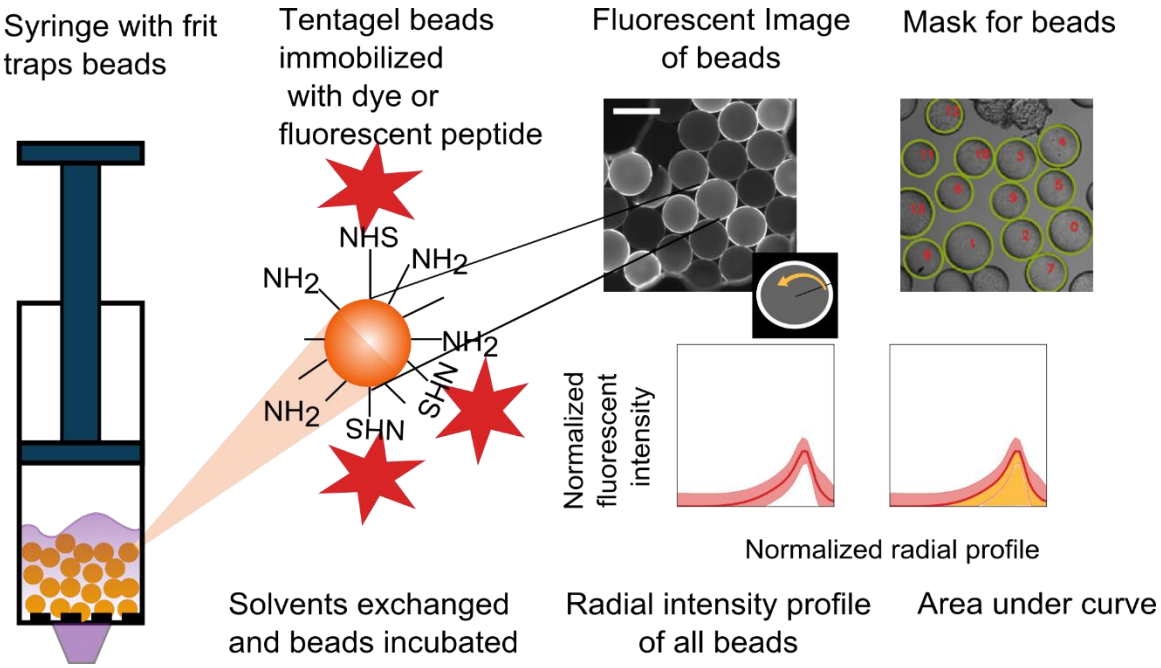

Figure 3.1: Specific binding of fluorophores to functionalized Tentagel beads occurs at the periphery and density can be measured by image processing.

The 100 μm amine functionalized Tentagel beads is incubated with the succinidimyl ester form of dye or peptide to form the stable amide bond. Repeated solvent washes remove the majority of non-specifically bound dyes or peptides resulting in abundant fluorescent signal at the bead periphery. A mask of every bead is generated and a radial intensity sweep for the fluorescent channel across each bead is performed. The radial intensity profile for a bead is normalized and shape corrected using a non-specifically bound dye on bead as a control. The area under the normalized radial intensity across all beads for an experiment is the density of the truly bound fluorophore or fluorescently labeled peptide on the bead. The scale bar shown in the fluorescent image is 200 μm.

RESULTS AND DISCUSSIONS

A small set of fluorophores are suitable for use with Edman solvents

Since the principle of fluorosequencing involves measuring the decrease in fluorescent intensity due only to Edman degradation experiment, it is critical that the fluorescence property of the fluorophores are not affected by incubation with solvents used in the chemistry (namely Trifluoroacetic acid (TFA) and pyridine).

The fluorophores, immobilized on Tentagel beads, were tested for changes in their fluorescence properties under prolonged 24 hour incubation at 40 °C with 9:1 v/v pyridine/PITC (reagent used for coupling reaction) and neat trifluoroacetic acid (reagent used for cleavage reaction) separately. Stability under these extreme conditions ascertains usefulness in shorter experimental cycles. The test on a palette of different classes of commercially available dyes spanning four excitation and emission filter spectra indicated that only a small number of fluorophores were suitable for the study. The fluorescence stability of the dyes after 24h TFA and PITC/pyridine incubation shortlisted six fluorophores that showed <40% change in fluorescence (see **figure 3.2a**).

Among the narrowed set of fluorophores in the red and far red fluorescence channels which showed a stable fluorescence, the dyes with rigid core structures such as rhodamine dyes (tetramethyl rhodamine, Alexa Fluor 555) and atto dyes (Atto647N) (shown in **figure 3.2b**) were used for further studies. Since the fluorescence imaging was performed at neutral pH, it is likely that the fluorescence properties of some of the chemically unstable fluorophores can be modified if the right protonation state is induced. Some dyes like Hilyte-488 and BODIPY-FL showed shifts in their fluorescence spectra after their incubation under acidic conditions and were incapable of reverting back to its original fluorescence profile after solvent washes and incubation with pH 7 buffer (see figure 3.2 for BODIPY-FL example).

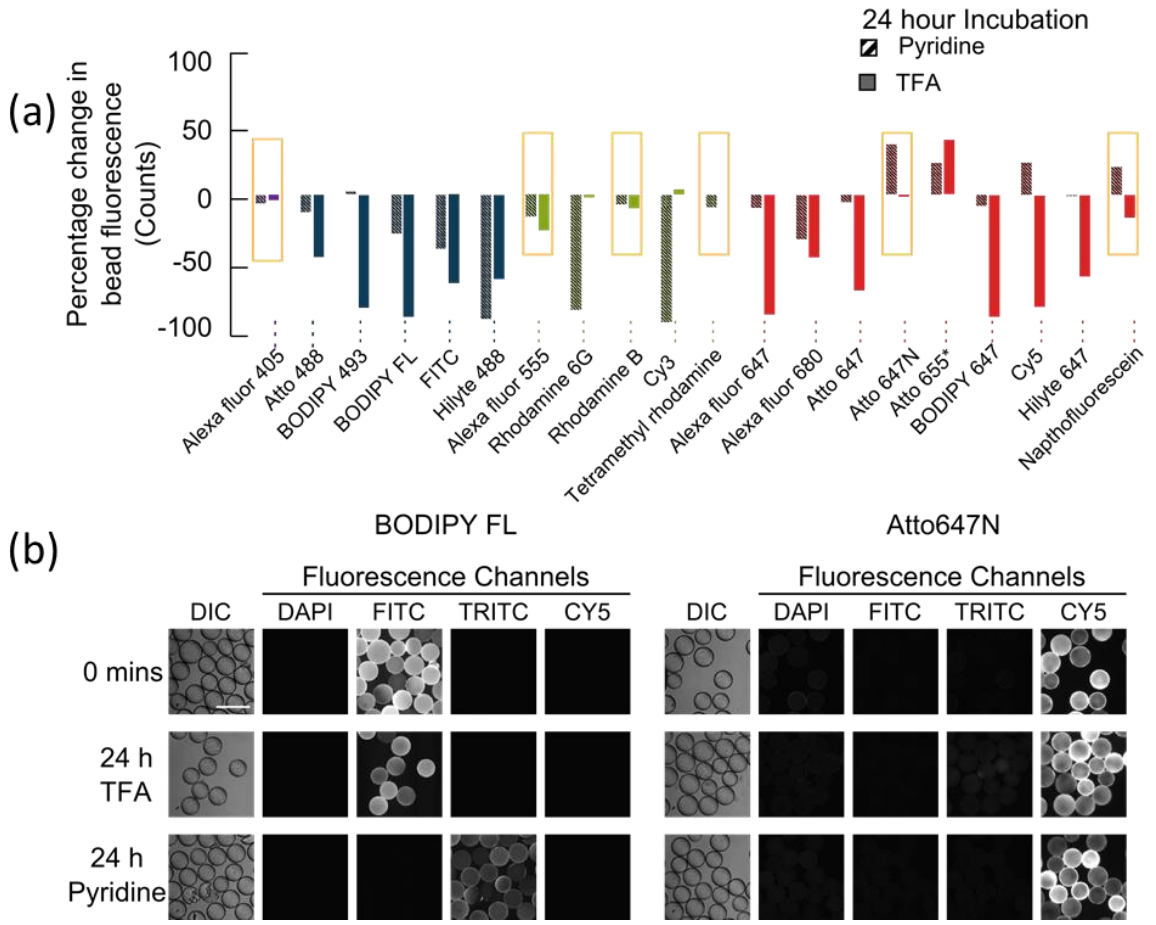

Figure 3.2: A select number of fluorophores exhibit fluorescence stability towards Edman solvents.

(a) The panel of fluorophores scanning across four fluorescent channels were tested for their percentage change in fluorescence intensity after a 24 hour incubation with trifluoroacetic acid (TFA) or pyridine/PITC in 9:1 (shown as pyridine) at 40 °C. The fluorophores demarcated in boxes had a relatively small change (<40%) in fluorescence with the prolonged incubation in the Edman solvents. (b) The panel of bead images are two examples of the fluorescence changes in dyes on the Tentagel beads with 24 hour TFA and pyridine incubation. The BODIPY-FL and Atto647N dye shows dramatic differences in dye behavior with the Edman solvent incubation. In the case of BODIPY-FL, the fluorescence intensity decreases with TFA incubation while there is a spectral shift with pyridine incubations. The fluorescence intensity is unchanged for the case of Atto647N dye. The terminologies used for the four fluorescence channels are combinations of filter sets described in methods section. The scale bar is 200μm.

While most of the dyes exhibited binding at the periphery, some fluorophores seemed to have high internal binding. Given the highly branched nature of the polystyrene bead matrix and the grafted polyethylene glycol layer, it is possible that the internal fluorescence represents non-specific binding of the dyes to hydrophobic pockets. Many fluorophores, which were added in large excess, could possess different extents of non-specific binding despite the repeated washes with solvents.

The reasons for the chemical instability of certain fluorophores are unclear and broad generalizations cannot be made based on core structure alone. Many commercially available fluorophores such as Hilyte647 (Anaspec, CA, USA) are packaged and sold with TFA salts and yet were not found to be acid stable under prolonged incubation. However, some empirical reasoning can explain the lack of stability of some fluorophores containing linear unsaturated bonds (polyenes), such as those found in cyanine or some BODIPY and Alexa Fluor dyes under prolonged TFA incubation. It is hypothesized that the protonation of unsaturated bonds under acidic conditions, induces a cis-trans isomerization reaction, thereby changing the underlying electronics of the fluorescence structure of the dyes [134]. Due to the commercial availability of cheap dyes and a long history on the study of rhodamine dyes and their functionalization, further studies involved rhodamine dyes, especially tetramethylrhodamine.

The amide bond formed between succinate ester and amine coated beads is specific and occurs at the bead periphery

The set of fluorophores stable to the Edman solvents also highlights the fact that the amide bond formed between the succinimidyl (succinate) ester of the fluorophores and the free amines on the Tentagel bead is chemically inert to the harsh Edman conditions used in the experiment. We tested the specificity of this amide bond formation by comparing it with control experiments involving a carboxyl or a hydrazide functional

group on Alexa Fluor 555 dye with the amine coated Tentagel beads (see **figure 3.3**). Internal binding of the dye was observed in these control experiments, while a clear peripheral binding was observed with the succinimidyl ester variant of the Alexa Fluor 555. We could not make any clear interpretation of the nature of binding between a maleimide variant of Alexa Fluor 555 with the thiol Tentagel beads and the isothiocyanate derivative of the tetramethylrhodamine dye due to its dispersed localization. This could also have been due to the poor loading of the fluorophore. The radial profile (shown in the image inset) elucidates the image processing methodology where covalently bound fluorophores are clustered in the periphery of the beads while non-specifically adsorbed fluorophores are trapped within the beads.

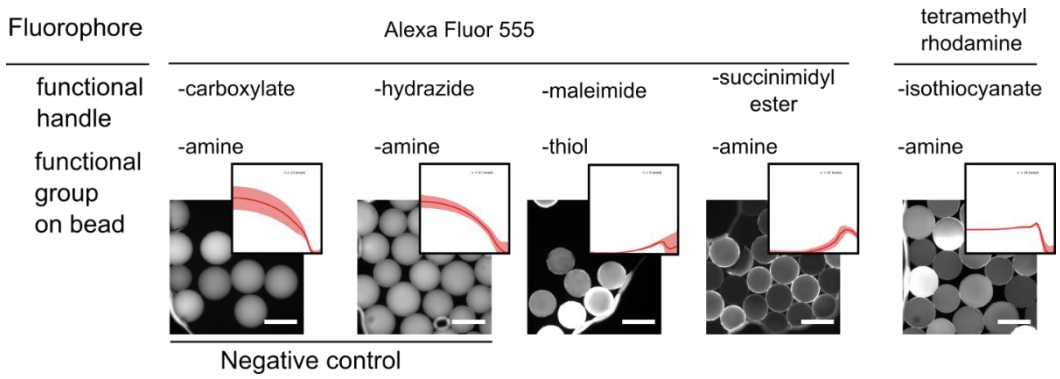

Figure 3.3: The amide bond between the dye succinimidyl ester group and the amine surface on Tentagel beads results in highly specific peripheral binding.

The panel of images shows the differences in binding profile (inset or radial distribution of fluorescence intensity in images) of the Alexa Fluor 555 with its different functional handles and tetramethylrhodamine isothiocyanate dye on amine or thiol grafted Tentagel beads. The negative control where the carboxylate and hydrazide derivatives of the Alexa Fluor 555 did not bind to the bead periphery but bound non-specifically in its interior. The binding nature of maleimide and isothiocyanate variants is unclear. Images are not contrast stretched and the scale bar is 200 μm.

Peptides can be covalently immobilized by their carboxyl functional group

Among the different immobilization schemes investigated, the knowledge of the stability of the amide bond between the succinate ester and amine surface was used to optimize a crosslinking procedure to immobilize peptides to the amine surface via their carboxyl termini [135]. Many solid phase Edman reactions have employed the use of EDC chemistry to immobilize peptides onto resin supports [85]. By performing EDC chemistry on amine coated glass beads and Tentagel beads, we have demonstrated a feasible scheme for covalently immobilizing peptides on the solid supports. It must be noted here, that the N-terminal amine group of the fluorescently labeled peptide is protected by either boc or fmoc protecting group that prevents the formation of the peptide concatemers. If the amines on the peptide are not protected, then amide bond formation would occur between the carboxyl and the free amine group of peptides in the presence of EDC. The fluorescence intensity of these immobilized peptides on Tentagel beads was unchanged with 24hour incubation with the Edman solvents (see **figure 3.4a**). Owing to the probable presence of hydrophobic pockets between the polymer matrices in Tentagel beads, which may give rise to false interpretation of binding, we performed the EDC test on aminosilane coated glass beads **(figure 3.4b)**. Under conditions prohibiting amide bond formation, we saw little to no binding on the glass beads. With these set of experiments, we have demonstrated a strategy to immobilize peptides covalently on amine surface and show the stability of the bonds and surface to incubations with Edman solvents.

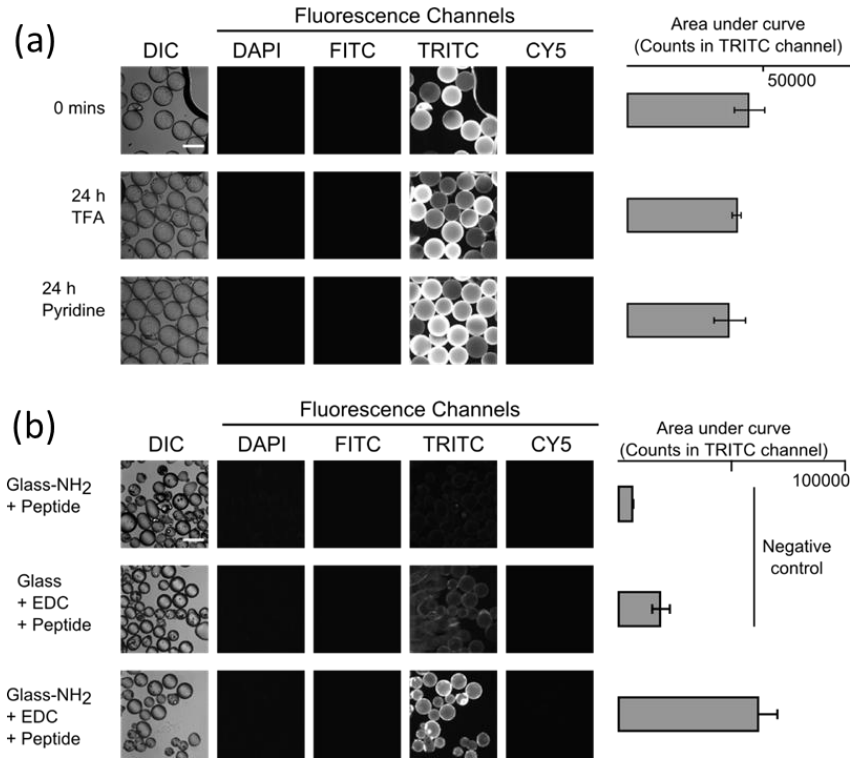

Figure 3.4: Peptides can be stably and covalently immobilized on amine surfaces using EDC chemistry.

The carbodiimide conjugation between the activated carboxylic acid of the peptide -(fmoc)-K*A (where * is the fluorescent tetramethylrhodamine) and amine group on Tentagel beads occurs by the EDC (Ethyl-(3-dimethylaminopropyl) carbodiimide)/NHS cross-linker (see methods for the EDC coupling protocol). The peripheral fluorescent signal from immobilized peptide is stable with 24 h incubation with TFA or pyridine/PITC (9:1 v/v) solvents as seen in figure (a). The panel of images of peptide (fmoc-K*A) immobilization on aminosilane coated glass beads in figure (b) controls for the effect of other variables involved in the chemistry and the non-covalent binding of Tentagel beads. High density peripheral binding of peptides is observed on amine coated glass beads which verifies that the peptides are covalently immobilized on the amine surface via their carboxylic acid group. Scale bar used is 200 μm.

Fluorescence of rhodamine dyes is pH dependent

The fluorescence from rhodamine dyes has been known to be pH dependent [136] requiring efforts to determine the most suitable imaging buffer. The investigation of pH dependence on the fluorescence properties of four different rhodamine labeled peptides

(see **figure 3.5** for structure and positional nomenclature for rhodamine dyes and the peptides), indicated an environmentally induced variation in their behavior.

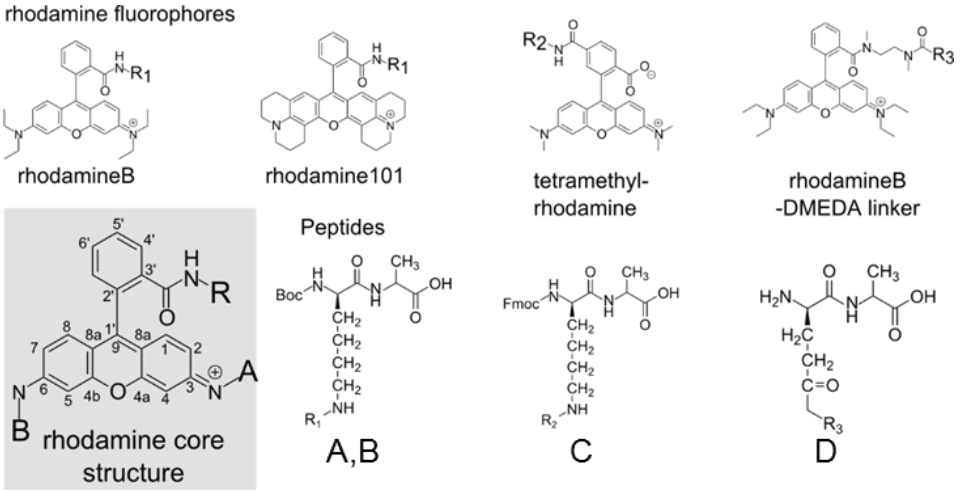

Figure 3.4: Structure of rhodamine variants with the conjugated peptide.

The structures of the four peptides with their ε–lysines labeled with the rhodamine dye variant are shown. Peptides A and B was (boc)-K*A labeled with rhodamineB and rhodamine101 respectively. Peptide C is (fmoc)-K*A labeled with tetramethylrhodamine and peptide D was labeled with a rhodamineB but contains a synthetic N, N'-dimethylethylenediamine (DMEDA) linker.

The acidic environment of the imaging buffer (pH 1.0) caused the highest fluorescence of the rhodamine labeled peptides (**figure 3.6a**). However the pH effect was most profound in the case of peptides labeled with rhodamineB (peptide A) and rhodamine101 (peptide B). This effect did not seem to occur for the case of tetramethylrhodamine labeled peptide (peptide C). The peptides A and B showed pH dependent fluorescence because the amide nitrogen foound at the 3' position is closer to the carbon position at 9 (or 1') and results in a 5 membered ring formation. This spirolactam ring is known to quench fluorescence and occurs at a pH higher than 3.1 [137]. This spirolactam formation does not occur for tetramethylrhodamine since the

succinate ester is present at the 5'- 6' position is not accessible to the central ring. The spironolactone formation, involving a ring formation with the carboxylate oxygen (at 3' position) can potentially quench fluorescence but requires a strong base such as piperidine. To test the hypothesis and prevent spirolactam formation in rhodamineB, we added an N, N'-dimethylethylenediamine (DMEDA) linker between the rhodamineB fluorophore and the aspartic acid side chain of the peptide resulting in the methylated amine at the 3' position. This prevented ring closure of the rhodamineB variant and was demonstrated by the independence of its fluorescence intensity with different pH imaging buffers.

By exploiting the fluorescence dependence on pH for the different fluorophores, we could isolate the fluorescence from a dye based on its pH and emission spectra. While the highest fluorescence of rhodamine B dye was observed in pH 1 buffer in the TRITC filter channel, the 5, 6-carboxynaphthofluorescein had its highest intensity in the pH 10 buffer in the Cy5 filter channel (**figure 3.5b**). This highlights a novel theoretical method of isolating two neighboring fluorophores from transferring resonance energy and thus preventing quenching or FRET (Forester Resonance Energy transfer) behavior [37]. The knowledge of the pH effect educated us on implementing pH 1 as the imaging buffer of choice for future experiments involving rhodamine dyes.

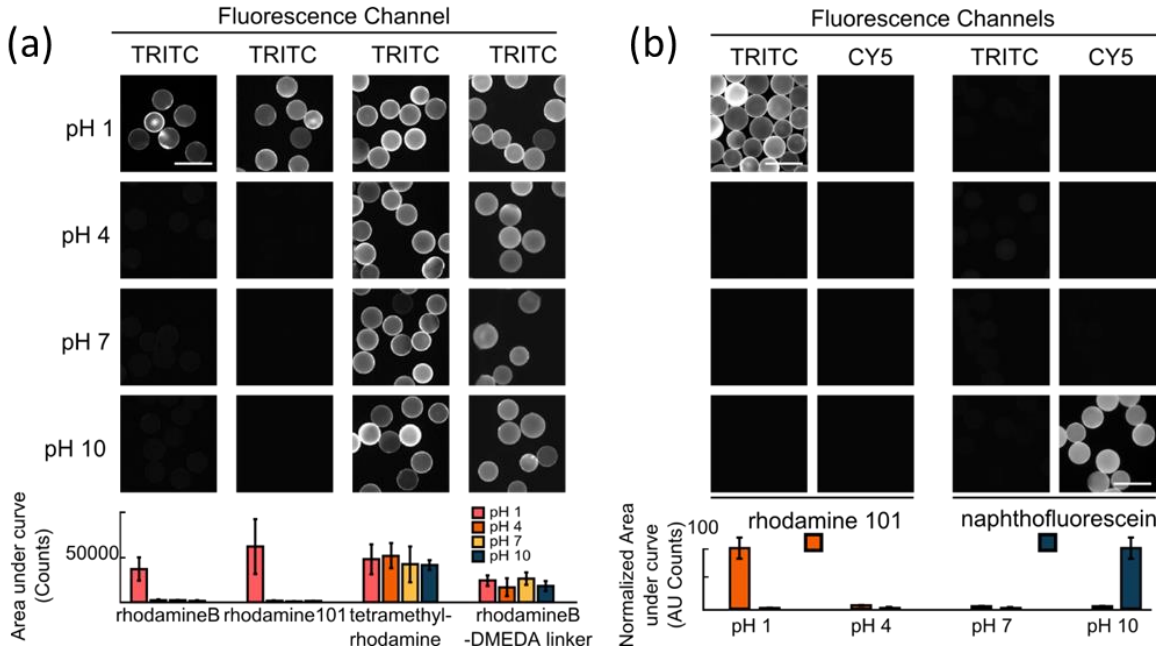

Figure 3.5: Fluorescence of rhodamine dyes attached to peptides is affected by the pH of the imaging buffer.

(a) Comparison of four different synthetic peptides (fmoc or boc)-K*A, labeled with commercially available rhodamineB, rhodamine101, tetramethylrhodamine or rhodamineB with DMEDA linker, showed differences in their fluorescence behavior under different pH conditions. While all the rhodamine variants showed enhanced fluorescence under pH 1 imaging buffer, the fluorescence of tetramethylrhodamine and synthesized methyl-rhodamineB was stable across pH 1 to pH 10 imaging buffers. The fluorescence of the rhodamine variants are dramatically reduced under basic conditions due to the formation of spirolactam (see text for reasoning). (b) The panel of images show the fluorescence of napthofluorescein in the CY5 channel under basic conditions while maximum fluorescence of rhodamine 101 dye is observed in acidic conditions. The pH effect on dye fluorescence can be theoretically leveraged to decouple the dye neighborhood interactions (such as FRET). The free electrons in the nitrogen atom in the amide bond formed with the peptide for rhodamineB and rhodamine 101 variants (see figure 3.5 for the structures) causes spirolactam ring formation and quenched fluorescence under basic conditions.

Edman degradation occurs at high efficiency on beads

After determining the stability of the fluorophore and the amide bond between the peptide's carboxyl and the surface's amine groups, we tested the efficiency of Edman chemistry on three different peptides differing in the position of its fluorescently labeled lysine residue. Four cycles of Edman degradation were performed in parallel on the three peptides with the sequences - (fmoc)-K*A, (fmoc)-GK*A and (fmoc)-K*AK*A (K* represents the lysine labeled with tetramethylrhodamine at its ε position). The peptides were immobilized on Tentagel beads via their C-termini and the fmoc protecting group at their N-termini was removed by incubation with 20% Piperidine in DMF for 1 hour prior to Edman degradation. To control for any false enhancements or decreases in fluorescence of beads due to effect of solvents and not the Edman chemistry, the "Mock" degradation scheme of solvent incubation and washes were used. A "Mock" Edman cycle is similar to a regular Edman cycle, but without the reactive phenylisothiocyanate reagent in the coupling solvent. The fluorescence profile of the beads through the Mock and Edman degradation cycles shows a statistically significant step drop coinciding with the position of the labeled lysine (**see figure 3.7**). Thus by tracking the fluorescence intensity decrease with Edman cycle, we obtain the positional information of lysine residues in the three peptides. The determination of this positional information is the basis for fluorosequencing.

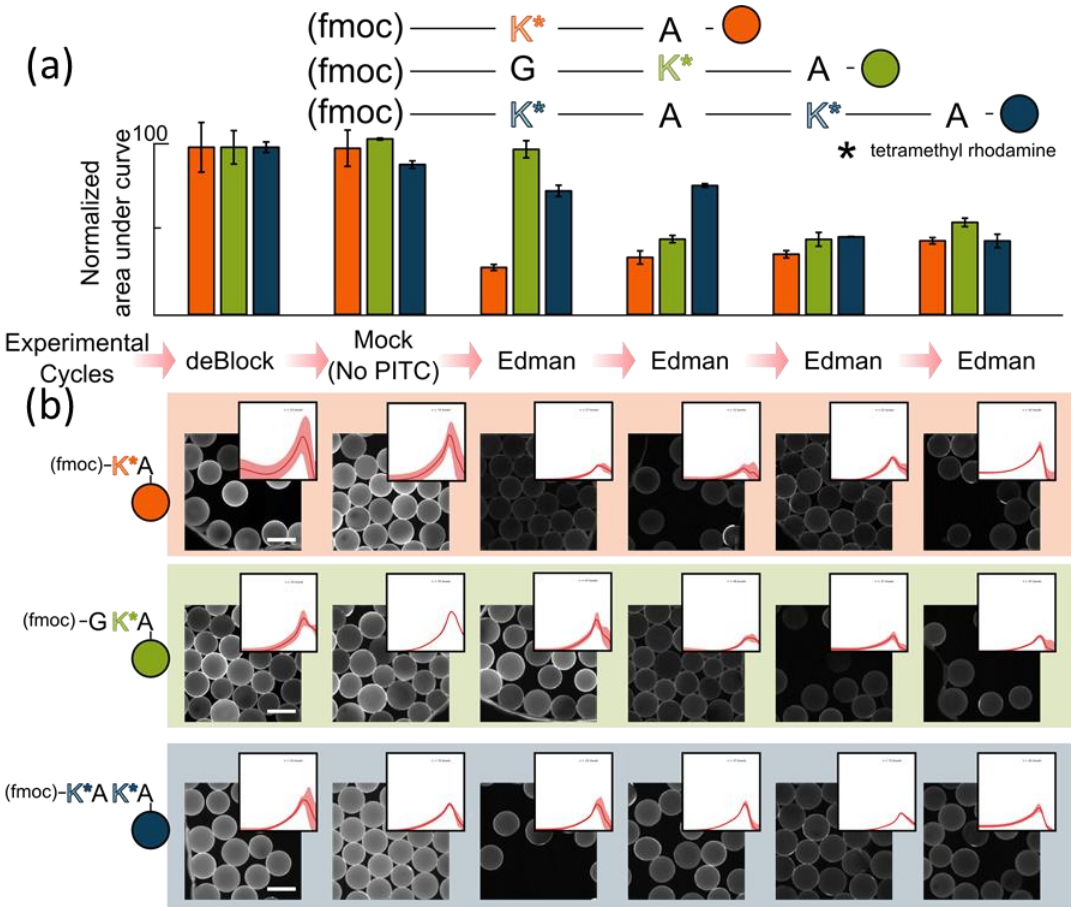

Figure 3.6: Edman degradation can be used to determine the positional information of the fluorescently labeled lysine residues of synthetic peptides using bulk fluorescence measurements.

Edman degradation was performed on three rhodamine labeled synthetic peptides (KA, GKA and KAKA) immobilized to Tentagel-NH2 beads *via* their C-terminal carboxyl group and blocked by fluorenylmethoxycarbonyl (fmoc) at their N-terminal amines. After deblocking the peptide, the step decrease in fluorescence intensity (in the TRITC channel) for each peptide coincided with the position of the labeled lysine as shown in the bar chart (a). Any loss of fluorescence occurring due to the use of solvents is controlled by the mock experimental cycle. A 60-70% decrease in the overall intensity after the Edman cycles is observed for all the beads. The panel of images (with the radial profile in the inset) are representative fluorescent images of the beads for each of the peptide used across all the experimental cycles. They provide a visual illustration of the decrease in the fluorescence of the beads that coincides with the position of the labeled lysine residue. All the fluorescent bead images are acquired in the TRITC channel (see methods for filter setup used) at an acquisition of 20 milliseconds under pH 1 imaging buffer. The scale bar shown is 200 μm.

The protocol used for Edman degradation was adapted and optimized from similar solid phase chemistry [70,78] and showed efficiency of cleavage ranging from 60-90%. It must be noted here that since Tentagel beads are heavily PEGylated (comprising of polyethylene glycol (PEG) polymers), a number of sites are available for strong non-specific binding of the hydrophobic peptides. Due to the accumulation of functional groups and thereby covalent peptide binding at the periphery of the bead we attempted to calculate the true fluorescence intensity of the peptides on the bead by calculating the area under its radial profile. Due to the unambiguous occurrence of a two-step drop in fluorescence intensity at Edman cycle 2 and 4 for the doubly labeled peptide (fmoc)-K*AK*A or the presence of a single step drop at Edman cycle 2 for the case of (fmoc)-GK*A, we could argue that the Edman efficiency must be largely greater than 50%, at least in the preceding steps. A lower efficiency would result in a decay of fluorescence with Edman cycles as opposed to a stepwise drop. The high efficiency of Edman degradation on these fluorescently labeled peptide variants establishes the practicality of performing fluorosequencing and Edman degradation on long fluorescently labeled peptides.

CONCLUSIONS

We have demonstrated the feasibility of identifying peptides immobilized on Tentagel beads in bulk, using the stepwise decrease in the fluorescence intensity with Edman degradation cycles. In order to achieve this - (a) fluorophores such as tetramethylrhodamine were analysed for their stability to solvents used in the chemistry; (b) the choice of an amide bond between the carboxylic group at the C-terminal amino acid of the peptide and amine groups on the surface by using EDC as the cross linker was

established; (c) the acidic buffer was identified as the ideal imaging solution and (d) the protocol for Edman degradation was optimized for high efficiency of amino acid cleavage.

Although this set of experiments provides evidence for the feasibility of fluorosequencing on the bulk fluorescence of peptides on beads, a number of key features of resin based chemistry must be considered when attempting to translate the chemistry to single molecule peptides immobilized on a flat glass surfaces. Some of them are - (a) The Tentagel beads were mixed vigorously with the Edman reagents during the chemistry; (b) The beads expand in certain solvents such as dichloromethane and could provide an enhanced local concentration of the reagents for the immobilized peptides; (c) Amine terminated polyethylene glycol chains, having a wide polydispersity, are grafted on the polystyrene beads and (d) High concentration of peptides were immobilized on the surface which may have resulted in high reaction rates. While it may be unclear on whether these factors can affect the translation to the alternate platforms, the establishment of the chemistry procedures (a prerequisite for single fluorosequencing studies) have been successfully established.

MATERIALS AND METHODS

Amine coating on beads

The commercially available 100 μm TentagelS-NH2 resin beads (Cat # 04773, Chem-Impex International Inc., IL, USA), made of amine functionalized PEG chains grafted on polystyrene beads, was used as such for the experiments. For the preparation of 100 μm glass beads (Cat # 4649, Sigma Aldrich, MO, USA) with an amine functionalized surface, the beads were loaded into syringe with frit (Cat # NC9214213,

Thermo Fisher) and first cleaned by repeated washes of 5% Alconox (detergent), followed by acetone, 90% Ethanol and finally 1 M Potassium hydroxide (KOH). Between each of the different solutions, the beads were thoroughly washed with de-ionized water. The aminosilane coating step was carried out by gently shaking the cleaned beads for 1h at room temperature in a solution of 10% Aminopropyltriethoxysilane (Cat # SIA0610.1 Gelest Inc., PA, USA) in the acidified 5% v/v of acetic acid/methanol solvent. The beads were washed with methanol and water before vacuum drying.

Peptides used in the study

The sequences and modifications of the custom peptides (provided by Dr. Eric Anslyn) are (a) (fmoc)-K[TMR]A, (b) (fmoc)-GK[TMR]A, (c) (boc)-K[rhodamine101]A, (d) (boc)-K[rhodamine B]A, (e) (boc)-K[rhodamineB-DMEDA]A and (f) (fmoc)-K[TMR]AK[TMR]A. Expansions of the abbreviations are - fmoc: fluorenylmethyloxycarbonyl, boc: butyloxycarbonyl, TMR: tetramethylrhodamine. The structures of the four rhodamine variants are shown in **figure 3.4**. All peptides were synthesized using a standard automated solid-phase peptide synthesizer, and purified using high-performance liquid chromatography (HPLC) or C_{18} solid phase extraction.

Peptide immobilization

For immobilizing peptide via the carboxyl group of the C-terminal amino acid, EDC chemistry [135] was used. About 40 nano-mole of the peptide, with the blocked amine at its N-terminal amino acid, was incubated with MES coupling buffer, comprising 6 mM EDC(1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride; Cat # 22980, Thermo Scientific), 5 mM NHS (N-hydroxysulfosuccinimide; Cat # 24599, Thermo Scientific) in 0.1 M MES buffer (pH 4.3; Cat #28390, Thermo Scientific), for 1h at room temperature. After appropriately diluting the activated peptides with 2 mM Sodium

bicarbonate buffer (pH 8.2; Cat # S233-3, Fisher Scientific), ~20 mg of amine functionalized beads were mixed and incubated for 16h at room temperature.

Fluorophore immobilization

The fluorophores used were either commercially purchased (from a number of distributors and vendors, predominantly Life Technologies, Sigma and Pierce) as a succinimidyl ester or chemically derivatized into that reactive form. The fluorophores, dissolved in dimethylformamide (DMF), were diluted in 2 mM Sodium bicarbonate solution (pH 8.2) to the appropriate concentration and incubated with Tentagel or glass beads for 16 hours prior to use.

Edman degradation procedure

The peptide functionalized beads were added into the syringes with frit, washed with DMF, dichloromethane (DCM) and methanol and dried under vacuum for 20 minutes. 20% Piperidine in DMF or 90% TFA in water was used to deprotect the fmoc or boc derivatized peptides respectively. In brief, the Edman reaction of the deprotected peptides on beads comprised of incubating the beads in 20% phenylisothiocyanate (v/v in pyridine) for 30 minutes at 40°C for the coupling condition, followed by incubating in TFA for 30 minutes at 40°C for the cleavage of the N-terminal amino acid from the peptide backbone. After the coupling and cleavage condition, the beads were washed with Ethyl acetate solution for 5 minutes with constant shaking. Following the Edman reaction and before imaging, the beads were washed thoroughly with DMF, DCM and methanol. All solvents used were reagent grade solvents purchased from Sigma Aldrich (MO, USA). For Mock experimental cycle, the entire Edman reaction was performed but PITC was not added to the coupling reagent.

Imaging of beads

A tiny portion (~0.5 mg) of the solvent washed and vacuum dried beads, which was added to 50 μL of pH 1 (0.1M KCl/HCl buffer) or other imaging buffers, was spotted on a clean glass slide. The beads were sandwiched with a coverslip and its sides were taped. The DIC and epi-fluorescence images of the beads were obtained using a Nikon Eclipse TE2000-E inverted microscope (Nikon Inc., Japan). The images of the beads were acquired at different exposure times with a Cascade II 512 camera (Photometrics, AZ, USA) on a Nikon Apo 10x/NA 0.45 objective. A combination of excitation filters DAPI – AT350/50 (340–380 nm), FITC – ET490/ 20 (465–495 nm), TRITC – ET555/25 (528–553 nm) and Cy5 - ET670 (590-640 nm) and emission filters DAPI – ET460/50 (435–485 nm), FITC – ET525/36 (515–555 nm), TRITC – ET605/52 (590–650 nm) or Cy5 – ET700/60 (640–730 nm) were used (Chroma Technology Corp, VT, USA). The use of corresponding excitation and emission filter set for the experiments described is represented by their filter name like DAPI, FITC, TRITC and Cy5 in the experiments. The Sutter Lambda 10-3 lter wheels (Sutter Instrument, CA, USA), motorized stage (Prior Scientific Inc. MA, USA) and image acquisition were driven by Nikon NIS Elements Imaging Software.

Image analysis of beads

For image processing and analysis, the circular outline of the beads was first identified by Hough algorithm. For a given fluorescent channels, the radial profile of every bead (normalized with its radius) was shape corrected with a negative bead profile (the radial profile of the control bead with only adsorbed fluorophores). This profile was averaged across all the beads under the experimental condition and area under the curve was calculated using the trapezoid method. For a different mode of image processing, when the peptide binding is not always on the periphery, masks were created for the

identified bead in the DIC channel and the count density (i.e. intensity/pixel) under the masks were calculated for all the fluorescent channels. All scripts were written in python using different publicly available image processing library such as openCV [138].

Chapter 4: Fluorosequencing of peptides at the single molecule level

INTRODUCTION

In the previous chapters we have discussed the theoretical justification for the fluorosequencing technique (chapter 2) and demonstrated the enablement of the fundamental chemistry procedures for peptide immobilization and Edman degradation using a bulk bead assay (chapter 3). Translating the bead based chemistry to a single molecule imaging system requires solving other fundamental and practical obstacles. This chapter contains a number of collaborative efforts in dealing with different aspects of the technique[6].

One of the major challenges in facilitating this single molecule fluorosequencing technique is the lack of any prior research in handling harsh organic solvents on microfluidic devices and performing fluorescence studies using it [44]. The use of harsh solvents such as trifluoroacetic acid limits the (a) number of inert materials that can be used in the system, (b) the choice of fluorophores and (c) the strategies for surface derivatization and subsequent peptide immobilization. Overcoming the technical challenges and implementing the fluorosequencing technique on synthetic peptide and peptide mixtures at the sensitivity of single molecules will demonstrate the proof-of-concept for the methodology.

The single molecule peptide sequencing system developed for this study required the integration of the fluidic device, a TIRF optical setup and a strategy for surface functionalization (**figure 4.1**). The fluidic system was automated and the Edman

[6] Alexander Boulgakov worked on the image analysis pipeline and provided a working program to analyze and summarize my image data. Erik Hernandez (from Dr. Eric Anslyn's group) synthesized a number of fluorescently labeled peptides. Joseph Marotta and Ken Demarest worked on automating and integrating the fluidic system with the Nikon NIS elements software.

chemistry procedure (developed for the bead study in Chapter 3) was implemented in the perfusion chamber. Algorithms for identifying and characterizing these single peptide molecules were also implemented.

In this chapter, I discuss the enablement of the setup to image single molecules, identify the position of fluorescently labeled amino acid residues in peptide sequences and provide evidence for discriminating peptides in a two-peptide mixture at a single molecule sensitivity, based on the fluorosequencing methods. Having solved some of the fundamental and technical challenges, it is conceivable to extend the technique to more complex peptides and protein mixtures.

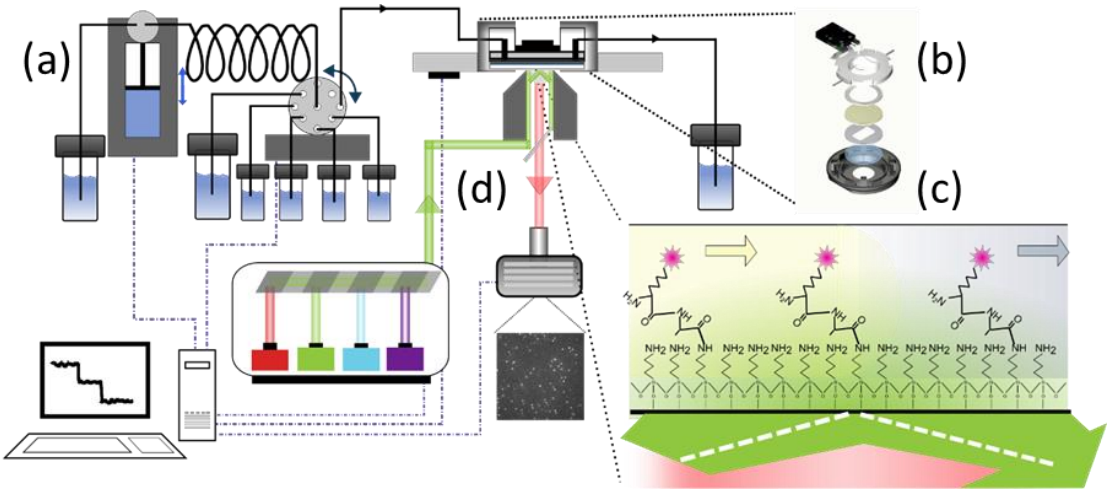

Figure 4.1: Schematic representation of the experimental setup.

The fluidic system (a) comprising of a syringe pump and a multi-position valve is automated to exchanges multiple solvents through the fluid line into the fluidic device mounted on the microscope platform. The exploded view of the fluidic device (image courtesy Bioptechs Inc.) (b) shows the internal parts of the temperature controlled chamber where fluid exchanges occurs over the functionalized glass surface. The fluorescence of the labeled peptides is acquired in a cooled EMCCD camera. The fluorescently labeled peptides are immobilized on the glass surface through their C-terminal end on an aminosilane surface (c). The optical setup (d) comprises the laser system from where the aligned laser beam illuminates the glass surface in TIRF mode.

RESULTS AND DISCUSSIONS

Peptides can be imaged at the single molecule level

By titrating in peptides or dyes on the amine functionalized glass surface with increasingly higher concentrations, we observed an increase in the number of bright fluorescent spots under TIRF illumination (with the appropriate laser and filter setup). In general, titrating in 2-20 pM peptide on an aminosilane functionalized glass surface for 1 hour results in 1000-2000 fluorescent spots per field (see **figure 4.2b**). Under constant illumination and high laser power, we can track the intensity of the identified fluorescent peaks through time. A number of known photo-physical processes characteristic of single fluorescent molecule such as blinking and photobleaching (by step drop in intensity) [46] were observed (see **figure 4.1c**). While the background fluorescence in the 647 channel was low and more or less stable, the high background fluorescence in the 561 channel decreased with illumination time (see methods for description of the incident laser and filter setup). The sum of the differences in the intensity of the central 3x3 pixel region from the mean intensity of the surrounding 16 pixel area (background pixels) was thus used to model the true intensity of the peptide. This decreasing background fluorescence (mostly observed in the 561 channel) is likely to be originating from fluorescing impurities used in the borosilicate glass for the coverslip. By rapidly traversing 1000 fields, with an overlap of 10% for the next field, a physical region of 4 mm (L) x 1 mm (W) on the glass surface was imaged. A total of 1.7 million peptides (K*EGAECGY; * fluorophore is Hilyte647) could be imaged in 2 hours using a single channel laser illumination and filter setup (**figure 4.2 a**). It must be noted that in the experiments involving single molecule fluorescence measurements, extreme precautions were taken to avoid the entry of any fluorescing impurities onto the imaging surface at every step of slide preparation and fluid exchanges. In all the experiments, imaging of the slide with

water and methanol was done before adding peptides or fluorophores to check for background fluorescence impurities. We have observed that the use of spectrophotometry grade solvents greatly diminish the occurrence of fluorescent impurities on the slide. The image acquisition and processing setup (see methods and figure 4.7 for the description) is capable of identifying and analyzing changes in the photophysics of single peptide molecules through time or experimental cycles.

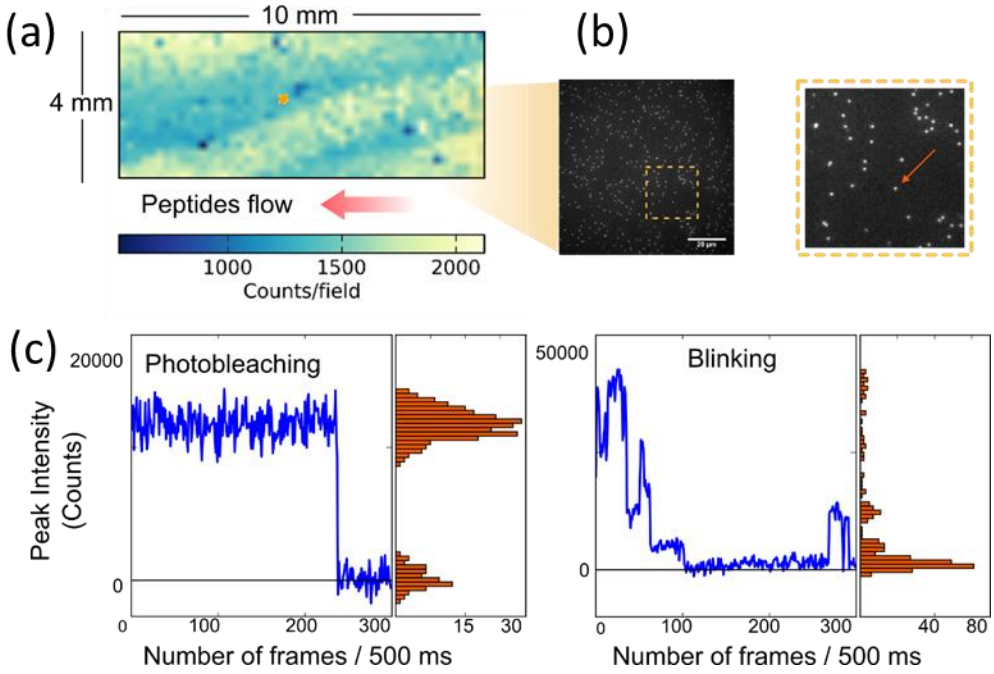

Figure 4.2: Imaging single peptide molecules.

(a) The heat map shows the density of immobilized peptides (K*[Hilyte647]EGAECGY) imaged across 1000 fields of view on the aminosilane glass surface which is housed in the microfluidic device. A total of 1.7 million peptides could be imaged in 2 hours using one imaging channel. (b) The typical field of view (dimensions of 200 μmx200 μm) in the 647 channel (peptide A: boc-GK [Atto647N] AGAG); see methods for the filter setup used) has an average of 1500 fluorescently labeled peptides. The annotation arrow of the fluorescent dot represents a single peptide molecule. (c)The intensity profile of the fluorescent signal under constant laser illumination exhibits two types of photo-physical phenomena of photobleaching and/or blinking characteristic of a single peptide molecule.

Degassed methanol is an ideal imaging solvent for reducing photobleaching

In the fluorosequencing technique, obtaining stable and high fluorescence signal from single peptide molecule is critical for statistically measuring intensity drops. Under different environmental conditions, the excited fluorophore may spend varying time in non-fluorescing states (e.g. triplet state) resulting in fluctuations in the observed fluorescence intensity from a single molecule called "blinking". Although the blinking phenomena plays an important part in localizing fluorophores as in super resolution microscopy [139], it is not ideal for determining the number of fluorophores on a molecule. It can be reasoned that if each fluorophore in a molecule blinked at different rates, the observed fluorescence emission would be a poor predictor of fluorophore counts. Minimizing the blinking behavior is thus important.

Step drops in fluorescence emissions occurring with a stochastic photobleaching events of a fluorophore has enabled counting the number of fluorophores per molecule and used in determining DNA repeats [125] or stoichiometry of proteins in a complex [45]. This step drops could be argued to be similar to fluorosequencing, where instead of the photobleaching event causing the destruction of the fluorophore, the cleavage event of the labeled amino acid by Edman degradation causes the step drop in fluorescence. Thus the imaging conditions, such as the time of acquisition, laser intensity and imaging buffers must be optimized to reduce the photobleaching and blinking rates.

While it may be essential in the study of single biological protein molecules [36] to use aqueous imaging buffers where the native structures are preserved, this is not a requirement for our experiments involving small fluorescent peptide molecules. Since methanol is used as the solvent of choice in organic chemistry to study photophysics of fluorophores [140,141], we decided to use methanol as our basic imaging solvent. Another advantage of using methanol is the enhanced solubility of triplet state quenchers

such as Trolox (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid) and Cyclooctatetraene [57,142] and thereby the ability to use these compounds at high concentrations in the imaging buffer. In addition, the alcohols such as methanol and ethanol, have a refractive index of 1.32 close to that of water (1.33). This ensures its adaptability to the TIRF microscopy setup.

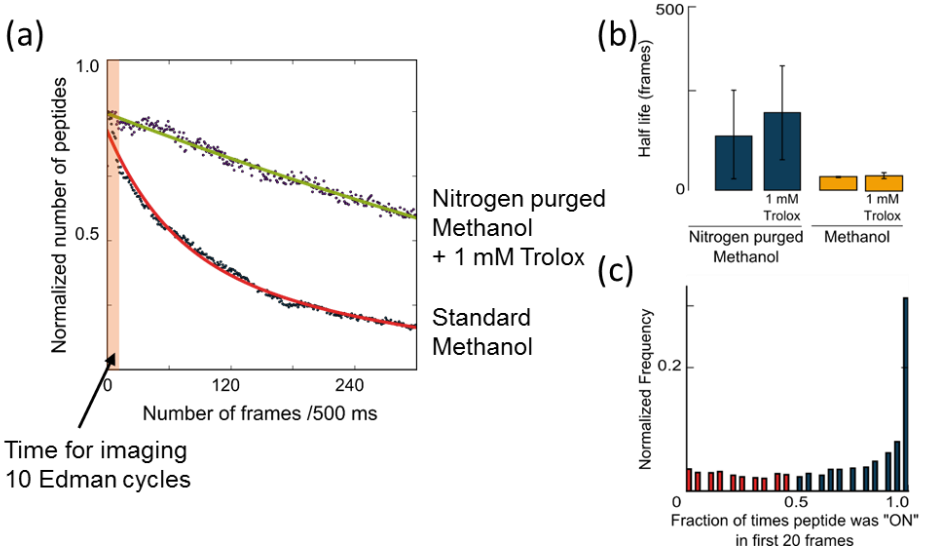

Figure 4.3: Nitrogen purged methanol with Trolox reduces photobleaching and blinking of fluorophores.

(a) The top green curve show the reduced time for photobleaching of tetramethylrhodamine labeled peptide, (fmoc)-K*A, under constant laser illumination when nitrogen purged methanol with 1 mM Trolox was used as the imaging buffer. (b) The half-life of the photobleaching curve under this imaging buffer was about 210 frames (or 105 seconds). (c) When this imaging buffer was used, the fraction of frames when the peptide was constantly fluorescing in the first 20 frames was about 30% as shown in the frequency histogram. The peptides were predominantly turned “ON” and did not blink or photobleach significantly during the imaging process.

In order to optimize the imaging conditions, we measured the intensity time traces of single molecule peptides ((fmoc)-K*A (*: tetramethylrhodamine)) under constant laser illumination for 5 minutes using different imaging buffers. An acquisition time of 500

milliseconds/frame allowed us to observe and analyze the phenomena of bleaching and blinking of the single molecules (see **figure 4.3**). In addition to the use of untreated methanol, we tested the effect of other imaging solvents, such as nitrogen purged methanol and the use of Trolox and cyclo-octatetraene as additives to methanol, on the rate of photobleaching and blinking. The combination of nitrogen purged methanol and 1 mM Trolox reduced the photobleaching effect by at least a factor of 5 (see **figure 4.3b**). It must be noted that the two step exponential fit best explained most photobleaching curves accounting for the rapid ground state depletion and the slower photobleaching phenomena [48]. However, under conditions with lowered photobleaching, such as with the addition of Trolox, the exponential fit was poor and thus the one step exponential fit model was used for measuring the photobleaching half-lives. To understand the effect of blinking rates, a normalized histogram of the dwell times (i.e. the number of frames that the molecule was identified as fluorescing) was plotted for the first 20 frames of acquired images. The 1 mM Trolox in nitrogen purged methanol had more than 30% peptides being constantly turned on when compared to the untreated methanol solvent (see **figure 4.3c**). It is however unclear, why the increase in the concentration of Trolox did not further decrease the photobleaching rates or blinking. Oxidized Trolox contaminants have been suggested to play a role in the photo-physical behavior of fluorophores [57]. These photobleaching studies illustrate the effect of oxygen removal from methanol and addition of triplet state quenchers.

Aminosilane coating on glass is stable to Edman degradation cycles

One of the key differences while translating the bulk fluorosequencing experiment performed on the Tentagel beads to the single molecule setup is the surface used for immobilizing the peptides. While the core of the bead consists of a polystyrene matrix,

single molecule TIRF experiments require fluorescent molecules to be immobilized on a #1 thickness glass coverslip. A widely used method for functionalizing glass and covalently immobilizing biomolecules is a self-assembled monolayer of aminosilane, consisting of an exposed aminopropyl group [60,143]. The simplest strategy for immobilizing fluorescent peptides or fluorophores to the glass surface requires the formation of two covalent bonds - the amide (bond between the succinate ester group of the fluorophore and the amine group) and the siloxane bond (Si-O-Si, bond formed between hydroxylated glass surface and silane layer) (see **figure 4.4a**). Although it has been shown in experiments on solid phase Edman degradation, that the aminosilane surface is chemically inert to the conditions and solvents used in Edman chemistry [86,144], acid hydrolysis of silane layers has also been reported [145–147]. Before we began the testing the fluorophores for trifluoroacetic acid stability (described in chapter 3), we had a strong hypothesis that the aminosilane layer was degrading with trifluoroacetic acid treatment. We then tested and characterized a number of different surfaces coatings, but by incorrectly using an acid susceptible dye (see **Appendix**).

The passivity of the immobilized peptides to solvent treatment relies on the stability of the amide and the siloxane bonds. The Tentagel bead assays (**figure 3.6**) confirmed the inertness of the amide bond to Edman solvents. To test the siloxane bond stability, we immobilized acid stable Atto647N fluorophore on the aminosilane layer via the amide bond and performed multiple cycles of Edman degradation. A constant density of peptides after the rounds of Edman cycles supported the hypothesis that the formed aminosilane coating is inert to Edman solvents and has specific chemical reactivity with exposed amine groups (see **figure 4.4b**). Having established the inertness of all the covalent bonds immobilizing the fluorophore to the surface, we were assured that

peptides could be stably anchored to the aminosilane surface by EDC chemistry and perform multiple cycles of Edman degradation.

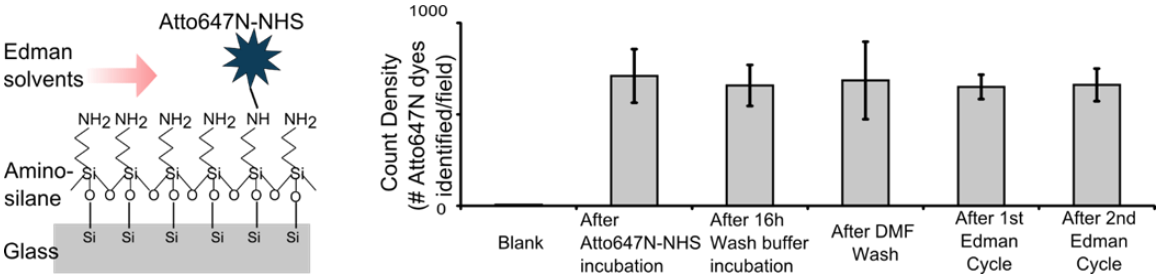

Figure 4.4: The aminosilane coating on glass slide is stable to Edman chemistry.

(a) The self-assembled monolayer of aminosilane (Aminopropyltriethoxysilane) is covalently grafted on a glass surface by siloxane bond formation (-Si-O-Si-). (b) The density of the Atto647N-NHS fluorophore, immobilized via a stable amide linkage to the silane layer, remains unchanged with repeated experimental cycles of Edman chemistry after washes with wash buffer (1% SDS and 0.1% Triton in 1X PBS buffer). This indicates that the silane layer is stable to the solvents used.

The Edman degradation cycle number provides positional information of the fluorescently labeled amino acid

Having integrated the component methodology such as the optical and fluidics setup, imaging conditions, surface coating and functionalization, image processing pipeline and Edman degradation chemistry on bulk peptides, we performed Edman degradation on fluorescently labeled peptide molecules using the single molecule setup. In a typical experiment, we first immobilize the N-terminal butyloxycarbonyl (boc) protected peptides to the aminosilane surface via their C-terminal end by EDC chemistry. After de-protecting the boc group by incubating it with 90% TFA (in water) for 5 hours, we perform two cycles of control 'Mock' chemistry. This Mock control step has the identical conditions and solvent exchange procedures as would be performed during the Edman chemistry cycle, but without the PITC reagent added to coupling solvent

(pyridine). These two cycles accounts for losses occurring due to removal of adsorbed peptide with the solvent exchanges and/or photobleaching. The Mock cycles are then followed by multiple experimental rounds of Edman chemistry. The same fields are imaged in methanol after the completion of each of these experimental cycles and the fluorescent intensity of single peptide molecules are tracked.

When performing the Edman cycles, the loss in fluorescence from a singly labeled peptide should coincide with the position of the dye labeled amino acid residue. We performed the multiple experimental cycles on the immobilized peptide, *Peptide A*: (boc)-GK*[Atto647N] AGAG, and observed a significant loss of about 50-60% in the average peptide density (images of 20 fields) only at the 2nd Edman cycle (**figure 4.5**). This coincides with the position of the Atto647N labeled lysine residue on the peptide and demonstrates that the loss of fluorescent peptides was predominantly due to the cleavage of amino acids by Edman chemistry. In order to verify that the loss of fluorescent peptides was indeed due Edman chemistry and not due to an unknown source of surface degradation occurring at the 2nd Edman cycle, we immobilized a spectrally distinguishable Alexa Fluor 555-NHS on the surface and tracked its density (shown as orange star). The constancy of the Alexa Fluor density through the experimental cycles, further confirmed that the loss of the fluorescent signal observed in the 647 channel was only due to the cleavage of the fluorescently labeled lysine residue by Edman chemistry. The panel of images of the same field shows the ability to track the fate of single peptide molecules through Edman cycles and identify the position of the fluorescent amino acid.

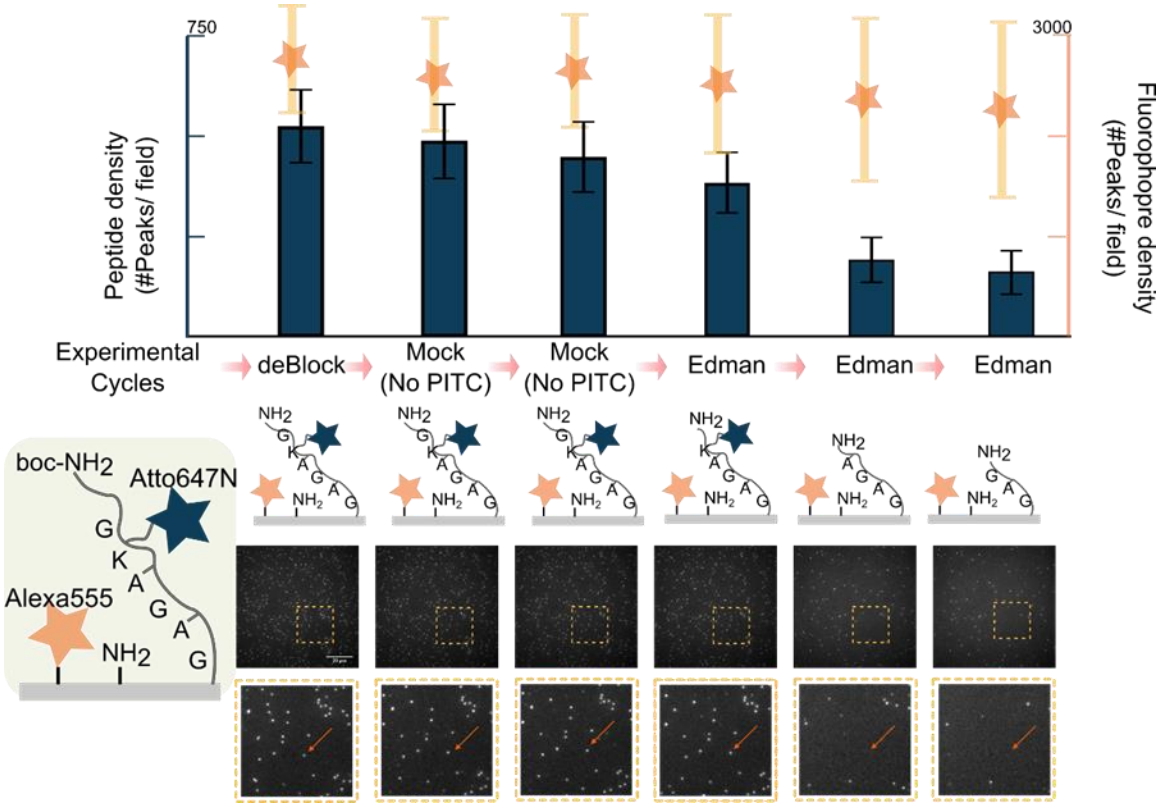

Figure 4.5: The position of the fluorescently labeled lysine can be determined by Edman degradation chemistry.

The density of fluorescently labeled peptides, (boc)-GK[Atto647N]AGAG, immobilized on the aminosilane surface, significantly decreased only at the 2nd Edman degradation cycle (as seen in the bar chart). This coincides with the position of the labeled lysine residue. There was no change in the density of the immobilized Alexa Fluor dye molecules (values with their standard deviation shown as orange stars) reaffirming the stability of the surface in the experiment. The panel of images are representative images where the same field through the experimental cycles is tracked, beginning with the removal of the boc group, two cycles of Mock chemistry (a control step where the reactive PITC reagent is removed) and three cycles of Edman degradation cycles. The pointed arrow indicates the observed disappearance of a single fluorescent peptide molecule only after the second Edman cycle.

We then tested the ability of this technique to discriminate two peptide populations differing in the position of their labeled amino acid residue. Two synthetic peptides – *Peptide A*: (boc)-GK*[Atto647N]AGAG and *Peptide B*: (boc)-

K*[Tetramethylrhodamine]AGAAG, differing in the position of lysine residue and the fluorophore (the two dyes fluoresces in the 647 and 561 channel respectively) were immobilized on the aminosilane surface. Experimental cycles of Mock and Edman chemistry was performed, as before, and the density of single molecules was measured in the two channels (see **figure 4.6**). The loss in the density of Peptide A (observed in the 647 channel) occurred only in the 2nd Edman cycle, that coincides with the position of its labeled lysine residue. A significant loss in the density of Peptide B (observed in the 561 channel) occurred at the 1st Edman cycle which correlates to the position of the labeled lysine residue in that peptide. The ability to assign the fluorescence signals to one of the two peptides at a level of single molecule (see image panel) in a mixture demonstrates the ability of the fluorosequencing technique to potentially discriminate peptide populations in a mixture.

The efficiency of Edman cleavage (say for Peptide A) is about 50% at the second cycle, if we calculate the loss of the peptide density from the preceding Edman cycle. We can speculate regarding the less than 100% efficiency of this reaction. Firstly the efficiency is clearly much higher than 50% in the preceding step of glycine cleavage, since a less than 50% decrease in peptide density in the 1st Edman cycle would have led to far fewer losses at the 2nd Edman cycle. In addition, there was not another step drop observed in the 3rd Edman cycle, indicating that further Edman cycles occurred with far less efficiency and/or peptides were unable to undergo further cleavage. The possibility of formation of an uncleavable peptide could be considered, since the presence of oxygen in the phenylhydantoin intermediate adduct results in oxidative desulfurization reaction [67]. This replacement of sulfur with oxygen results in a urea bond (phenylcarbamyl group) which is incapable of undergoing further Edman reactions. This could also account for the fact that cleavage does not occur with further Edman cycles. In other

words, whichever peptide could have undergone cleavage, did undergo Edman degradation.

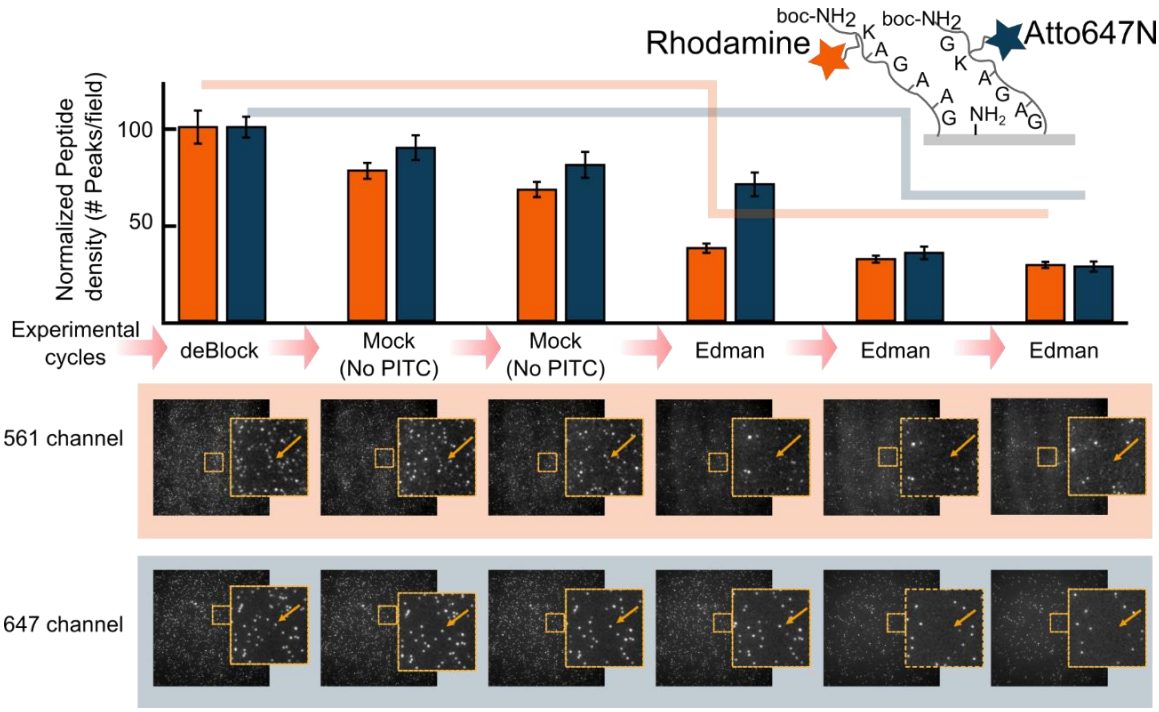

Figure 4.6: Peptides differing in their fluorescently labeled amino acid position can be discriminated at the level of a single molecule.

The fluorescent density of two spectrally distinguishable peptides immobilized on the aminosilane surface, decreased significantly at the first and the second Edman degradation cycles and was observed in the 561 and 647 channels respectively (as shown by the bar charts). While the decrease in the fluorescent peptide count in the 561 channel can be ascribed to the loss of rhodamine labeled lysine in the peptide – (boc)-K*AGAAG, the loss of peptides in the 647 channel corresponds to the Atto647N labeled lysine cleavage in the peptide – (boc)-GK*AGAG. Single peptide molecules were tracked through the experimental cycles of de-protection, two control Mock cycles (solvent exchanges without the PITC reagent) and three Edman degradation cycles (as shown in the panel of images). The arrows annotating the fate of two spectrally different peptide molecule through Edman cycles, is an example demonstrating the ability to discriminate two population of peptides in a mixture at a sensitivity of a single molecule.

CONCLUSIONS

In this chapter, we have demonstrated the technique to identify the position of a fluorescently labeled amino acid residue in peptides at a sensitivity of a single molecule. The technique involved optimization of the Edman chemistry, adaptation and automation of a fluidic system and a set of image acquisition and image processing methods to enable the fluorosequencing on the single peptide molecules. We have been able to obtain a working solution for a number of practical obstacles such as screening solvent compatible materials, imaging surfaces and fluorophores. We have shown the ability of the fluorosequencing technique to discriminate peptides in a mixture with high sensitivity based on its position of its fluorescently labeled lysine. The technique could have implications in large scale multiplexed protein identification.

MATERIALS AND METHODS

Aminosilane slide coating

40 mm #1 thick glass coverslips (Bioptechs Inc., PA, USA), was placed vertically in a custom made Teflon rack, and cleaned by washes and sonication with 5% Alconox (detergent), acetone, 90% Ethanol and finally 1 M Potassium hydroxide (KOH). Between each of the different solvent washes, the slides were thoroughly washed with de-ionized water. The aminosilane coating step was carried out by incubating the slides for 20 minutes in 1% Aminopropyltriethoxy silane (Cat # SIA0610, Gelest Inc., PA, USA) dissolved in the acidified 5% v/v of acetic acid/methanol solvent. The slides were sonicated intermittently for 1 minute to dislodge any adsorbed silane molecules. After incubation, the slides were rinsed thoroughly with methanol and water. It was then dried with nitrogen and stored under vacuum until use. The slides were imaged in water and

methanol prior to peptide or fluorophore immobilization to check for presence of fluorescing impurities.

Peptides used

Peptide A: (boc)-GK*[Atto647N]AGAG and Peptide B: (boc)-K*[Tetramethylrhodamine]AGAAG was synthesized by Thermo Fisher Scientific (IL, USA) with a purity of >95% and validated by mass spectrometry. Peptide K[Hilyte647]EGAECGY was synthesized by AnaSpec Inc. (CA, USA) and validated by mass spectrometry. Peptide (fmoc)-K [Tetramethylrhodamine]A was synthesized and purified by solid phase HPLC by Dr. Eric Anslyn's group. In all the peptides, the fluorophores was covalently attached to the ε-amine of the lysine residue

Peptide immobilization

400 picomoles of peptide (dissolved in 20 μL dimethylformamide) was incubated with 80 μL of 0.2 μm filtered MES coupling solution (0.1 M MES buffer (Cat #28390, Thermo Scientific), 5 mM N-hydroxysuccinimde (NHS; Cat # 24599, Thermo Scientific) and 6 mM (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride) (EDC; Cat # 22980, Thermo Scientific) for 1 hour to form an EDC adduct at the carboxyl terminus of the peptide. Dilutions of this EDC coupled peptide was made to zepto-molar concentration in 0.2 μm filtered 2 mM sodium bicarbonate buffer (pH 8.2; Cat # S233-3, Fisher Scientific). The peptides were titrated on the slide and imaged at every dilution to reach an approximate density of about 1000 peptides/field. The peptides at the appropriate dilution were then incubated for 1 hour at room temperature for immobilizing the peptide to the amine functionalized surface via the amide bond. Non-specifically bound peptides were removed by repeated washes with water and methanol.

Solvents used

Highest purity and mostly spectrophotometry grade solvents of Methanol (Cat # 494437, Sigma), Ethylacetate (Cat # 270989, Sigma), Acetonitrile (Cat # 34967, Sigma), trifluoroacetic acid (Cat # T6508, Sigma), Pyridine (Cat # 270970, Sigma), Dimethylformamide (DMF, Cat # 270547, Sigma), phenylisothiocyanate (PITC, Cat # P1034-10x1ml, Sigma) and water (Cat # 5140, Thermo Scientific) was used for all the experiments. Coupling solvent, comprising of 9:1 v/v of pyridine: PITC, was freshly prepared before use. The coupling solvent and the free-basing solvent consisting of 10:3:2:1 v/v of acetonitrile: pyridine: triethylamine: water was flushed with nitrogen for 5 minutes and maintained under nitrogen atmosphere by piercing the septum with a nitrogen filled balloon. The cleavage solvent used was 90% TFA in water. The glass vials fitted with a sealable Teflon-silicone septum (Cat # 27022, Sigma) used was rinsed with acetone and the solvent with which it is stored. The FEP tubing from the valves were pierced through the septum and the entire system was maintained under anoxic condition.

Fluidics system

The aminosilane coated glass coverslip housed in a microfluidic chamber was adapted from the FCS2 perfusion chamber (Bioptechs Inc., PA, USA). The vendor supplied upper and the lower gaskets was replaced with 0.03" perfluoroelastomer Kalrez®-0040 material (DuPont Inc., local vendor - Austin Seals company, TX, USA) and a diamond shape was cut in the lower gasket (die Number - 452458, cut by Bioptechs Inc.). The shape ensured complete fluid exchanges when compared with a rectangular cut. The Kalrez material had ideal compressibility with a shore durometer A of 70 (personal communication with supplier) and had chemical inertness to trifluoroacetic acid (**figure 4.7**).

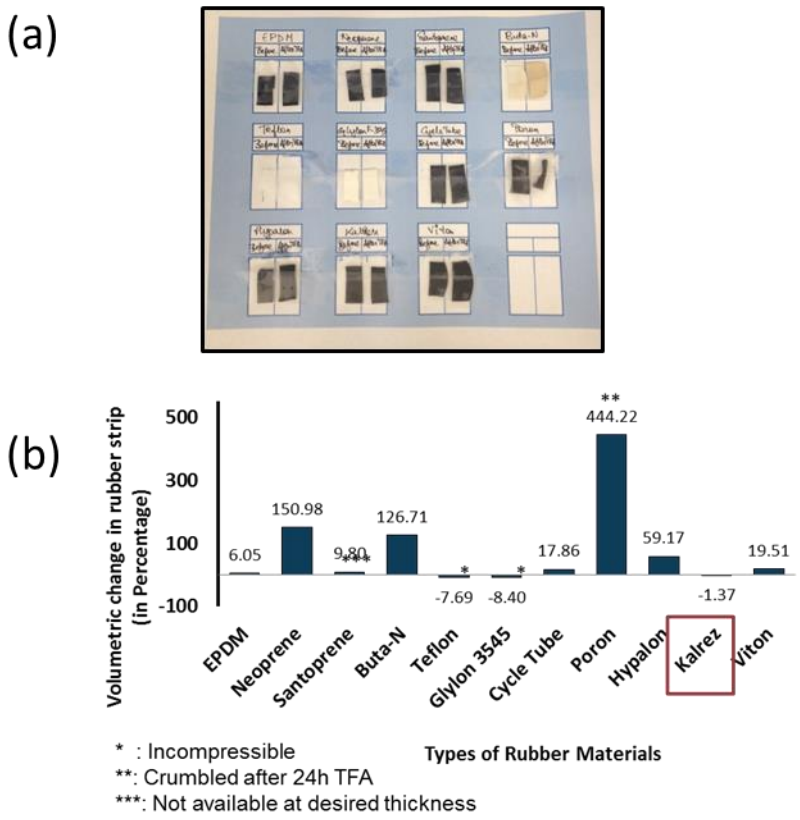

Figure 4.7: Kalrez® rubber is the ideal gasket material for handling trifluoroacetic acid

(a) 11 different rubber materials sourced through a number of vendors were cut as 2cm (L) x 1cm (W) strips and tested for its inertness after 24 hour of TFA incubation (see the panel of cut strips, before and after TFA incubation). (b) Among the tested rubber materials, Kalrez ® (from DuPont) material showed the least change in volume (as seen in the bar chart) and had a shore durometer A of 70, indicating good compressibility. Teflon although inert was not compressible and caused leaks when used in the perfusion chamber. Other materials such as Santoprene was not available at the desired thickness.

The fluid flows through the perfusion chamber was controlled by a pump and valve system. This sequential injection system, comprising of a 1 mL 3-port syringe pump (Cat # MicroCSP-3000) and 10-port multi-position valve (Cat # C25-3180EMH) was used to dispense distinct solvents used in Edman chemistry into the perfusion chamber. Water was used as the carrier fluid and the tubing were flushed before every experiment. This sequential injection system was purchased from FIALab Inc. (WA,

USA). 1/16” fluorinated ethylene propylene (FEP) tubing was used in all the connections leading to and from the pump and valve system. A union connector (Cat # P-623, IDEX Health and Sciences, WA, USA) combined the 1/16” tubing from the valve output to the 1/8“ tubing connecting to the perfusion chamber inlet. The outlet from the perfusion chamber was a 1/8“tubing draining into a waste container.

The pumps and valve was controlled via its RS-232 serial port which was integrated with the computer by a RS232 to USB converter. Custom made code in python (using pySerial module) was used to control the speed, position and volume of fluid dispensed or aspirated by the pump and the different open positions of the valve via the vendor’s built in firmware. A set of sequential commands to the pump and valve automated the Edman procedure.

Imaging system

The inverted Nikon Ti-E (Nikon Inc., Japan) microscope equipped with a perfect focus system and a motorized stage - ProScan II (Prior Scientific, MA, USA) has a custom built insert (from Bioptechs Inc., PA, USA) to hold the microfluidic device. The Agilent laser launch unit - monolithic laser combiner (MLC400B) has 405, 488, 561 and 640 nm wavelength solid state lasers coupled to the AOTF crystal with a nominal power output of 20 mW for the channels. The output laser is coupled by a fiber optic cable to the inverted microscope via the TIRF arm, which contains a focusing knob to align the light beam onto the back focal plane of the 60X 1.45 oil objective. The 561 (Cat # 97323) and 647 (Cat # 97336) channel TIRF cube (Nikon Inc., Japan) has the appropriate cleanup emission filter and excitation filters. Images acquired using this 561 filter cube and the 561 incident laser is termed as images in the 561 channel and similarly 647 channel if the 647 filter cube and 647 incident laser combination is used. The images are

recorded with an iXon-X3 DU897 EMCCD camera (Andor Inc., Belfast, UK) cooled to -70C at a typical exposure of 1 sec per frame. The lasers, camera and stage movements are controlled by NIS Elements (Nikon Inc., Japan) and/or by the macros scripted into the software environment.

Image analysis

The image processing pipeline for identifying and characterizing the fluorescent signal from a single molecule is shown in **figure 4.8**.

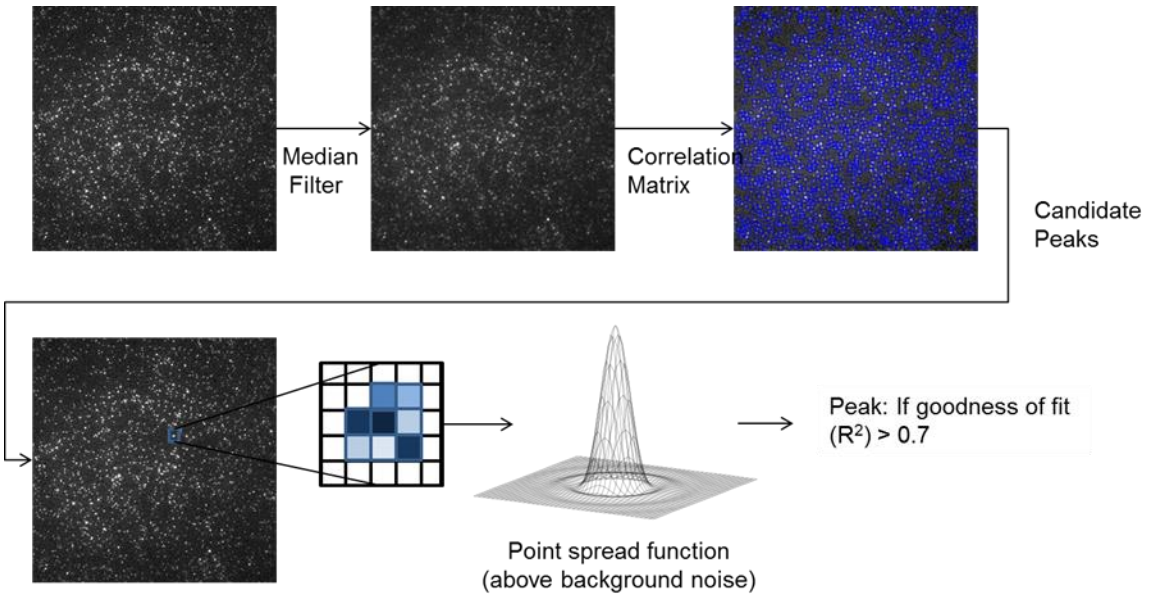

Figure 4.8: Image processing pipeline to identify single molecules

Identification and attribution of fluorescent signals to single molecules is first achieved by applying a median filter to the acquired images followed by candidate peak selection using the convolution algorithm (see methods for details on the correlation matrix used). The pixel intensities (spanning a 25 pixel area) of candidate peak regions are fitted with a Gaussian function to approximate to a point spread function. The peak is assigned to single molecule signal if it passes a goodness of fit criteria (R2>0.7).

In detail, a 512x512 pixel image from the camera was losslessly converted to tiff and png files. Median filter was applied on the png image to subtract the background. A convolution algorithm (with the values for the 9x9 convolution matrix populated by averaging the intensity values of 2000 manually identified peaks) was used to detect candidate peaks (which could be a fluorescing molecule) within a 5x5 pixel area. A Gaussian fit (since it approximates the airy function) was attempted on these candidate peaks to model a point spread function. Poorly fitting peaks with an R^2 value of the fit less than 0.7 was discarded from the analysis. If the distance separating two adjacent peaks is less than 3 pixels, the peak with the lower R^2 fit value is discarded from analysis. The sum of the central 3x3 pixel area was used as the estimate of the peak intensity and the ratio of the deviation of the intensity from the peripheral background 16 pixels to the deviation of the background pixels was used as the metric of signal to noise ratio (SNR). Alignment of two images of the same field on the slide was performed by rigid body registration algorithm and the image fields were typically offset by about 2-5 pixels in the x and y direction. No significant rotation of the images was detected.

Imaging condition

While most of the experiments involved untreated but spectrophotometry grade methanol (Cat # UN1230, EMD Millipore, Germany) as the imaging solvents, other solvents were used to optimize the imaging conditions. 100 μM, 1 mM and 10 mM concentrations of Trolox (Cat # 238813, Sigma) was prepared in methanol and added to the untreated methanol or nitrogen purged methanol. The imaging solvents were allowed to equilibrate with the immobilized peptides on the slide for at least 30 minutes before imaging.

Chapter 5: Conclusions and future perspectives

This dissertation discusses one of the first attempts in developing a working system to identify single peptide molecules in a mixed peptide population. We first developed a theoretical underpinning for the fluorosequencing technology which helped in understanding the potential of the method and also ideate working solutions to account for the various foreseen errors. For the implementation of the fluorosequencing technique, we had to first establish the principal procedures for peptide immobilization and Edman degradation on fluorescently labeled peptides. To achieve this, we developed a bead based platform to screen for fluorophores that were stable to the harsh Edman solvents and optimize the chemistry protocols, in order to demonstrate fluorosequencing in bulk. Having established the Edman chemistry and found the fluorophores compatible with the solvents used, we implemented the technique to sequence simple synthetic peptide molecules using a single molecule setup. By developing this setup, a working solution to a number of fundamental and practical obstacles was established, such as the material choice for gaskets, purity of solvents used for the chemistry, imaging processing algorithms and an imaging buffer capable of reducing photobleaching and blinking phenomena. By identifying the fluorescent position of labeled lysine residue in a synthetic peptide and in a two-peptide mixtures, we demonstrated the feasibility of the fluorosequencing technique for single molecule peptide sequencing.

Although a number of the component technologies used in the fluorosequencing technique had precedence, the integration into a new setup, produced a number of challenges to overcome. One of the major challenges faced during the project was the initial difficulty in interpreting a decrease in peptide density with trifluoroacetic acid

(TFA) incubation. We had made a series of reasoning that the fluorophore, Hilyte647, was inert to acid incubations. The assumptions were due to the fact that (a) the synthetic dye labeled peptide requires TFA to cleave it from resin (b) the dyes are sold as the TFA salt (personal communication with supplier) and (c) some of our early experiments of incubating the dye with TFA and the removal of TFA did result in the fluorescence recovery. The project led us on a time consuming path in trying to develop an alternate surface to aminosilane. However, the discovery that the fluorophore is susceptible to acid treatment proved useful in propelling the technique forward. Another challenge in the project was the need to screen a number of rubber materials, for gasket use, that is inert to TFA. Understanding the solvent effects on dyes was found to be a crucial step in designing the final experimental protocols. We finally had to solve the problem of obtaining a clean surface without fluorescent impurities, especially in the 561 channel. This required screening highly pure solvents which were used for slide preparation, Edman chemistry and imaging studies.

The short term goal in advancing the fluorosequencing technology would be to establish the fundamental question in the effect of dye-dye interactions on fluorescence. Sequencing synthetic peptides containing (a) two labels of the same type, such as two rhodamine fluorophores, and (b) two spectrally distinguishable fluorophores (such as 5,6 -carboxynapthofluorescein as the 1st dye and rhodamine variant as the 2nd dye) would aid in understanding the fluorophore interactions such as FRET or quenching. This may even require developing interesting design strategies of fluorophore structures in order to minimize spectral overlaps. The environmental effects on fluorescence discussed in the thesis, could be a novel work around in selectively decoupling the dye-dye interactions.

Sequencing insulin (which comprises two chains of 30 and 21 amino acids) will be a bridge towards sequencing more complex protein mixtures. Insulin is relatively

inexpensive and available with high purity. The insulin protein molecule has two chains and provides an internally built stoichiometry. This means that when sequencing the peptides at the single molecule level, we should be able to identify the two peptide chains in equal proportions. The presence of five cysteine residues (3 and 2 respectively in the 2 chains) would provide an easy experimental setup to label the peptides with available fluorophore derivatives (Alexa647N-Iodoacetamide (Sigma Aldrich) or the carboxynapthofluorescein dye with the iodoacetamide handles (synthesized by Dr. Eric Anslyn's group)). By using GluC to digest insulin (the oxidized insulin B chain is the model peptide used to validate the GluC quality control test by Sigma Aldrich), we can obtain peptide fragments that differ in the position of cysteine residues (3 fragments contain 1 cysteine and 1 fragment contains 3 cysteine residues) and can be used for identifying the four peptide chains. This would extend the proof of the fluorosequencing technique to sequence naturally occurring proteins.

What are the prospects for turning this technique into a practical protein sequencing technology? With a 60 year old heavily optimized and well understood Edman chemistry procedure, along with a number of commercially available Edman sequenators, it is easy to envision the improvement of our current 50-60% cleavage efficiency. Our experiments are already highly parallel and offer an unprecedented sensitivity in identifying peptide sequences in a mixture. Using our current imaging setup, we were able to identify about 1.5 million peptides in a 40 mm^2 area. It is conceivable to scale the imaging area to a 25x25 cm region and thereby make it feasible to sequence at least a billion peptide molecules. This may however require redesigning the imaging platform in conjunction with a perfusion chamber. Some engineering efforts would also be required to develop imaging systems that can rapidly acquire images across multiple channels. This is important because a 1 second exposure per field

(currently done in the thesis), translates to 280 hours (or 11 days) for imaging 1 billion peptide molecules for one experimental cycle. There is however precedence for solving many of these obstacles, as similar fluidic and imaging needs have been met by the Helicos (now a defunct company) or Illumina DNA sequencing platforms. The image processing algorithms developed for the project is scalable to handle large quantities of data and graphical processing units, present in most modern computers, can speed up computational time. We envision constant improvements for removing systematic biases in our experimental protocols, build more internal controls and optimize image analysis pipelines. This technology has the potential for sequencing proteomes ranging from simple viruses to complex cancers.

Appendix: Surface chemistry

INTRODUCTION

Based on an incorrect hypothesis that trifluoroacetic acid (TFA) treatment is capable of acid hydrolysis of silane layers, we attempted a number of different coating alternatives on glass surfaces (Table A1). The choice of surface coating was primarily driven by three factors – surface (a) must be coated with a thickness of less than 50 nm (b) have a refractive index close to water to aid alignment in the existing TIRF setup and (c) must have organic functional groups (such as amines or carboxylates) or be capable of being functionalized. We characterized a large number of these surfaces and tested the stability of (acid susceptible) fluorescent dye to trifluoroacetic acid incubation. Among the different surfaces coated on glass, a fluoropolymer, Cytop (Bellex Inc., USA) had a number of interesting properties and was nearly suitable for the fluorosequencing technique. In this appendix, I describe experiments on working with this polymer for TIRF microscopy and the attempts in functionalizing the coated surface[7].

[7] The functionalization scheme described here was developed by Joseph Marotta. I characterized the surface. Hugo Celio (Center for Nanosciences Facility) helped me in running the XPS experiment and analyzing the spectra.

Table A.1: Survey and characterization of surfaces for TIRF microscopy

Surface coating	Deposition method	Functionalization method	Surface characterization
No coating; Bare glass	Ammonia plasma Allylamine plasma	Amine functionalized monomers deposited	Visibility: Dye observed; XPS: small loss of allylamine peaks; Si peaks observed after TFA incubations
Aminosilane	Liquid phase; Gas phase; PECVD	Plasma deposition; Variants of aminosilane used; Repeated cycles with TFA incubation and Aminopropyltriethoxy silane (APTES) deposition; Mix of fluorosilane and APTES	Dyes observed; Loss of dye density with TFA incubation. No change in contact angle for fluorosilane; Small change in dipodal silane; XPS: Small loss in carbon peak.
Teflon AF	Spin coating 1% of polymer in Fluorinert (FC-72)	Ammonia plasma (PECVD)	Dyes and beads were visible but surface was uneven. No change in contact angle after TFA treatment
Gold *Cr(10 nm);Au(50 nm) *Ti(5 nm); Au (50 nm)	Thermal deposition; E-beam deposition	Thiol-PEG-functional group added	Unobservable dyes, peptides or quantum dots
Polystyrene *Polystyrene *Poly(styrene-b-tertbutylacrylate)	Spin coating; Melting polystyrene beads (220°C for 20 minutes)	Melting of amine functionalized beads; TFA hydrolysis	High amount of fluorescent impurities on melted beads; XPS data shows speculative results of carboxylic acid formation
Polymethyl-methacrylate sheet	Melting on glass at 110°C for 30 minutes	Plasma treatment for chemisorption	Surface was opaque with heptane wash

RESULTS AND DISCUSSIONS

Thin coating of CytopM on glass can be used for TIRF microscopy

Cytop is an amorphous fluoropolymers from Asahi Glass (Japan; Sold through Bellex International, Wilmington DE,USA) which exhibits superior chemical resistance, 95% transparency to visible light and has a refractive index of 1.34 (Technical notes from Bellex International, [148]). A variant of Cytop, Cytop M, (see **figure A.1** for structure) has an amino-silane coupling group to the end polymer. This promotes the adhesion of Cytop M to glass without the need for applying a primer adhesive (such as fluorinated silane) to glass. A 1% and 2% Cytop-M solution in CytopSolv, was spin coated, at different speeds, on the 40 mm glass coverslips and 2" Silicon wafers. The coating on silicon wafer acted as a proxy for measuring thickness of the Cytop layer by Ellipsometry (see **figure A.2a**). With a speed of 4000 rpm a uniform Cytop surface with a 20 nm thickness was prepared.

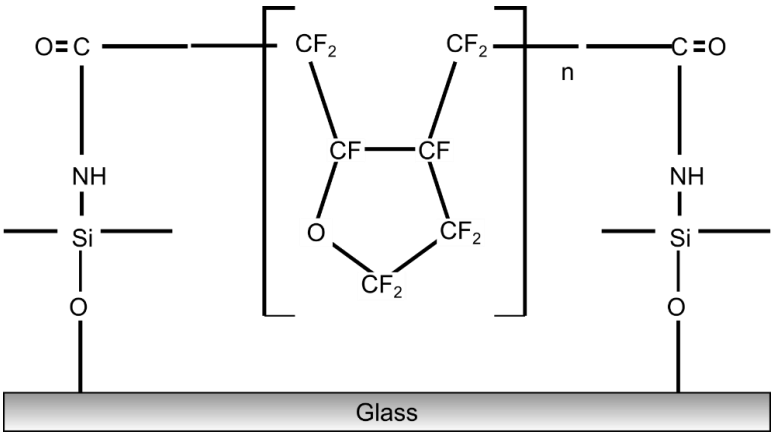

Figure A.1: Structure of functionalized CytopM

The chemical structure of the amorphous fluoropolymer CytopM is shown. In this variant of Cytop, the end group is functionalized with aminosilane to improve adhesion to glass substrates or oxidized substrates.

Atomic force microscopy measurements were performed on the Cytop coated surface and it had a roughness of only about 1.2 ± 0.7 nm. (**Figure A.2b**). The polymer coating and curing process helped form a smooth surface. When I incubated the slide with fluorescent peptide molecules (K*[Hilyte647] EGAECGY), random movements of these molecules were observed. A snapshot of the movie is shown in **figure A.2c**. This indicated that the Cytop surface provides a low background surface that can be used for TIRF microscopy.

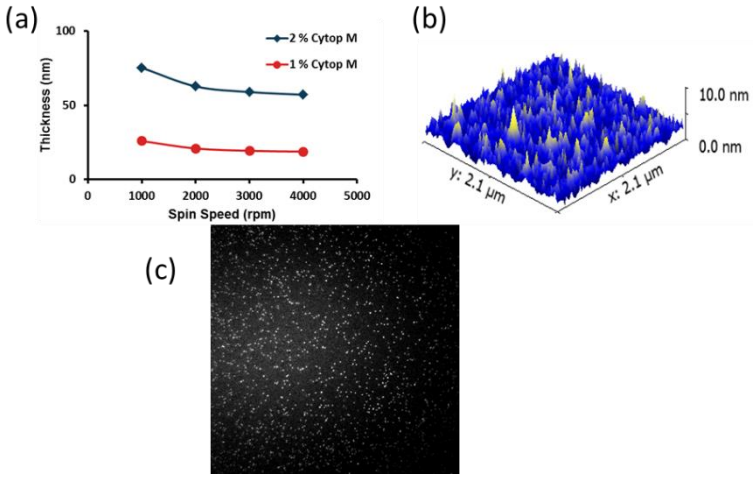

Figure A.2: Thin CytopM layer can be spin coated on a glass surfaces

(a) 1% and 2% CytopM was spin coated at different speed on a cleaned Si substrate. Ellipsometry (technique used to measure thickness of coating layers) measurements indicated than 20 nm thick CytopM surface can be obtained, making it amenable to TIRF illumination. The Si substrate was used as a proxy for glass measurements since the technique cannot work on the glass substrate. (b) Atomic Force Microscopy (AFM) studies on the CytopM layer on glass coverslip measured a roughness measurement of about 1.5 nm. (c) Image shows the high signal to noise of the peptide (K*[Hilyte647]EGAECGY) added to the Cytop surface (no immobilization). It also indicates that Cytop can be used for TIRF imaging.

Cytop coating is inert to TFA incubation

The Cytop coating produces a hydrophobic coat with a water contact angle of about 110°. A 24 h incubation of the glass slide coated with Cytop with TFA did not alter the contact angle significantly (see **figure A.3b**). The figure insert shows the water droplet on the Cytop surface. We also characterized the surface with X-ray photoelectron spectroscopy, a widely used method to characterize the nature of the bonds on surfaces (see **figure A.3**). The survey spectra of CytopM remained unchanged after TFA incubation. On the region scan of the C1s, O1s and F1s with a dwell time of 120 seconds, characteristic peaks representing the different types of Carbon-Fluorine bonds in Cytop was observed. However no noticeable change in the peaks was observed with 24 hour TFA incubation. These two independent lines of evidence confirm the highly passive nature of the Cytop coating.

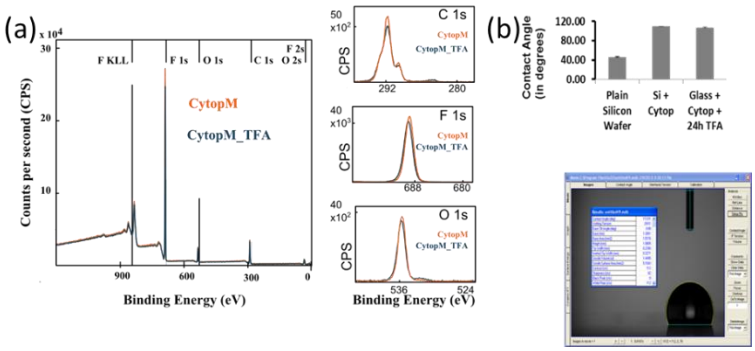

Figure A.3: Surface characterization of CytopM coating shows stability to TFA incubation

(a) Surface coating of 1% CytopM on glass was characterized by X-ray photoelectron spectroscopy (XPS; a technique that provides information of the chemical bonds present on the surface, like C-C or C-O etc.) for changes in surface bonds with TFA incubation. By comparing the two spectra (orange and blue denoting Before and After 24h TFA treated surfaces), no noticeable change was observed in the survey spectra. The three regional spectra analyzing the C 1s, F 1s and O 1s bonds also did not show any significant differences. (b) Contact angle measurements of Cytop surface with water also did not indicate any difference with TFA treatment.

Functionalization of CytopM

After a number of attempts in functionalizing the surface by ammonia plasma, allylamine grafting and even doping functionalized fluoropolymer during the Cytop coating process, we did not seem to form stable dye binding with TFA treatment. A procedure of introducing carbonyl group was adapted from an earlier study attempting to improve the wettability of Cytop [149]. Aluminum was thermally deposited on Cytop coated glass and its removal by NaOH incubation resulted in a hydrophilic surface[8]. XPS characterization of the surface (see **figure A.4**) indicated two additional peaks in the C1s and O1s region scans. This had a lower binding energy. By aligning the peaks based on known carbon peaks, adventitious carbon and the spectral shift (made with respect to Si 2s peaks) and XPS databases (http://xpssimplified.com), we could speculate the formation of carbonyl like bonds appearing on the surface.

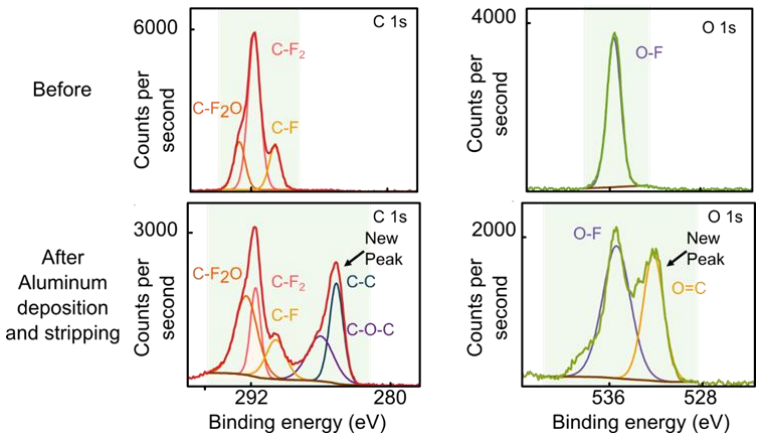

Figure A.4: The Aluminum deposition and stripping method for functionalizing CytopM resulted in carbon and oxygen bond formation.

The XPS spectra in the C1s and O1s region of CytopM layer (before and after the Aluminum functionalization method) showed a new peak formed at lower binding energy to the parent peak after the functionalization method (see annotation of new peak). The peaks were assigned to carbon-oxygen bonds.

[8] Joseph Marotta performed the entire aluminum functionalization experiment.

CONCLUSIONS

We repurposed a fluoropolymer material, Cytop for TIRF microscopy and demonstrated its inertness to trifluoroacetic acid. We used a method for introducing functional groups and by characterizing the surface, we speculated the formation of carbonyl-like groups exposed on the coating. However the different batches of Cytop and/or the functionalization process introduced a large number of fluorescent impurities and thus making it difficult to perform further TIRF microscopy.

MATERIALS AND METHODS

Coating Cytop layer

To achieve a low thickness (<50 nm) coating, Cytop M concentrations and spin speed of the spin coater was optimized. The spin condition of 400 rpm for 1sec and a slowdown to 0 rpm in 10 seconds was used for coating Cytop using spin coater in the clean room facility in Center for Nanosciences (Speciality Coating Systems, TX, USA). The glass coverslip is then degassed at 60°C for 10 minutes. This is followed by ramping the temperature to 160°C over 30 minutes. The surface was then cured at 160°C for 1 hour.

Measurement of thickness by Ellipsometry

A 2" Silicon wafer (111) (TedPella, CA, USA) was used as a proxy for the glass coverslip for measuring thickness of the Cytop coating using Ellipsometry (M2000 Spectroscopic Ellipsometer, J.A. Wollam, NE, USA). Measurements were made at five different angles between 41-71 degrees in the wavelength range of 300-1000 nm. The Cauchy equation for soft material was fitted on to the experimental data to obtain the thickness. The measurements were made in triplicates for each sample.

Surface Roughness measurements by Atomic Force Microscopy

The topography of the Cytop coated glass slides were studied by tapping mode Atomic force microscopy (Agilent 5500 Atomic Force Microscope, Agilent Technologies, CA, USA; Facility in Center for Nanosciences) using a silicon cantilever with a conical shaped etched probe, with a resonant frequency of 160 KHz and a spring constant of 5 N/m was used (MikroMasch, CA, USA). The laser was aligned on the silicon tip and the reflected beam was centered on a QPD (Quadrant Photo - Diode) detector. The deflection of the tip caused during scanning the surface features is measured and topography mapped via a Proportional-Integral feedback control. The operating parameters typically used for surface scanning are – (a) proportional and integral gains were adjusted between 5 – 20% (b) scan speed was between 0.5 – 1 line/sec (c) resolution of 256x256 (d) Scan area was typically 1 - 10 μm^2. All surfaces were dried by nitrogen before analysis and the surface scanning performed in technical triplicates. The data was analyzed by Gwyddion 2.31 software and the roughness of the surface was calculated at five different regions on the image.

Surface compositional analysis by X-ray photoelectron microscopy

The elemental composition and the nature of the bonds on the different treated surfaces were analyzed by X-ray photoelectron spectroscopy (XPS). The data was recorded using the monochromatic Al-K_α line at 1486 eV and 150W and a pass energy of 80 eV and 20 eV for acquiring the survey spectrum and highly resolved elemental spectra (like C 1s, Si 2p, F 1s etc.). Charge neutralization was implemented at the default setting of 20 mA. Data acquisition and processing were carried out using Vision software (Kratos, Manchester, UK) and CasaXPS version 2.3.16 (Casa Software Ltd, Teignmouth, UK) respectively. Peaks were fitted with Gaussing-Lorentzian peaks after linear subtraction of the background.

Bibliography

1. Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature. Nature Publishing Group; 2003;422: 198–207. doi:10.1038/nature01511

2. Blackstock WP, Weir MP. Proteomics: quantitative and physical mapping of cellular proteins. Trends Biotechnol. 1999;17: 121–127. doi:10.1016/S0167-7799(98)01245-1

3. Shen PS, Park J, Qin Y, Li X, Parsawar K, Larson MH, et al. Rqc2p and 60S ribosomal subunits mediate mRNA-independent elongation of nascent chains. Science (80-). 2015;347: 75–78. doi:10.1126/science.1259724

4. Maier T, Güell M, Serrano L. Correlation of mRNA and protein in complex biological samples. FEBS Lett. 2009;583: 3966–73. doi:10.1016/j.febslet.2009.10.036

5. Weston AD, Hood L. Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. J Proteome Res. 2004;3: 179–96.

6. O'Sullivan BP, Freedman SD. Cystic fibrosis. Lancet. 2009;373: 1891–904. doi:10.1016/S0140-6736(09)60327-5

7. Walker FO. Huntington's disease. Lancet. 2007;369: 218–28. doi:10.1016/S0140-6736(07)60111-1

8. Hanash S, Taguchi A. The grand challenge to decipher the cancer proteome. Nat Rev Cancer. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2010;10: 652–60. doi:10.1038/nrc2918

9. Ning M, Lopez M, Cao J, Buonanno FS, Lo EH. Application of proteomics to cerebrovascular disease. Electrophoresis. 2012;33: 3582–97. doi:10.1002/elps.201200481

10. Zhang J, Keene CD, Pan C, Montine KS, Montine TJ. Proteomics of human neurodegenerative diseases. J Neuropathol Exp Neurol. 2008;67: 923–32. doi:10.1097/NEN.0b013e318187a832

11. List EO, Berryman DE, Bower B, Sackmann-Sala L, Gosney E, Ding J, et al. The use of proteomics to study infectious diseases. Infect Disord Drug Targets. 2008;8: 31–45.

12. Rusling JF, Kumar C V, Gutkind JS, Patel V. Measurement of biomarker proteins for point-of-care early detection and monitoring of cancer. Analyst. 2010;135: 2496–511. doi:10.1039/c0an00204f

13. Verrills NM. Clinical proteomics: present and future prospects. Clin Biochem Rev. 2006;27: 99–116.

14. Craig R, Cortens JP, Beavis RC. Open source system for analyzing, validating, and storing protein identification data. J Proteome Res. American Chemical Society; 2004;3: 1234–42. doi:10.1021/pr049882h

15. Wang M, Weiss M, Simonovic M, Haertinger G, Schrimpf SP, Hengartner MO, et al. PaxDb, a database of protein abundance averages across all three domains of life. Mol Cell Proteomics. 2012;11: 492–500. doi:10.1074/mcp.O111.014704

16. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, et al. Towards a knowledge-based Human Protein Atlas. Nat Biotechnol. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2010;28: 1248–50. doi:10.1038/nbt1210-1248

17. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science (80-). 2015;347: 1260419–1260419. doi:10.1126/science.1260419

18. Kodadek T. Protein microarrays: prospects and problems. Chem Biol. 2001;8: 105–115. doi:10.1016/S1074-5521(00)90067-X

19. Barry R, Soloviev M. Quantitative protein profiling using antibody arrays. Proteomics. 2004;4: 3717–26. doi:10.1002/pmic.200300877

20. Alhamdani MS, Schröder C, Hoheisel JD. Oncoproteomic profiling with antibody microarrays. Genome Med. 2009;1: 68. doi:10.1186/gm68

21. Xie F, Liu T, Qian W-J, Petyuk VA, Smith RD. Liquid chromatography-mass spectrometry-based quantitative proteomics. J Biol Chem. 2011;286: 25443–9. doi:10.1074/jbc.R110.199703

22. Wang K, Huang C, Nice E. Recent advances in proteomics: towards the human proteome. Biomed Chromatogr. 2014;28: 848–57. doi:10.1002/bmc.3157

23. Kim M-S, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, et al. A draft map of the human proteome. Nature. Nature Publishing Group; 2014;509: 575–81. doi:10.1038/nature13302

24. Wilhelm M, Schlegl J, Hahne H, Moghaddas Gholami A, Lieberenz M, Savitski MM, et al. Mass-spectrometry-based draft of the human proteome. Nature. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2014;509: 582–7. doi:10.1038/nature13319

25. Havugimana PC, Hart GT, Nepusz T, Yang H, Turinsky AL, Li Z, et al. A census of human soluble protein complexes. Cell. Elsevier; 2012;150: 1068–81. doi:10.1016/j.cell.2012.08.011

26. Thakur SS, Geiger T, Chatterjee B, Bandilla P, Fröhlich F, Cox J, et al. Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. Mol Cell Proteomics. 2011;10: M110.003699. doi:10.1074/mcp.M110.003699

27. Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, et al. Deep proteome and transcriptome mapping of a human cancer cell line. Mol Syst Biol. EMBO and Macmillan Publishers Limited; 2011;7: 548. doi:10.1038/msb.2011.81

28. Ghaemmaghami S, Huh W-K, Bower K, Howson RW, Belle A, Dephoure N, et al. Global analysis of protein expression in yeast. Nature. 2003;425: 737–41. doi:10.1038/nature02046

29. Nivala J, Marks DB, Akeson M. Unfoldase-mediated protein translocation through an α-hemolysin nanopore. Nat Biotechnol. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2013;31: 247–50. doi:10.1038/nbt.2503

30. Zhao Y, Ashcroft B, Zhang P, Liu H, Sen S, Song W, et al. Single-molecule spectroscopy of amino acids and peptides by recognition tunnelling. Nat Nanotechnol. Nature Publishing Group; 2014;9: 466–473. doi:10.1038/nnano.2014.54

31. Ohshiro T, Tsutsui M, Yokota K, Furuhashi M, Taniguchi M, Kawai T. Detection of post-translational modifications in single peptides using electron tunnelling currents. Nat Nanotechnol. 2014; doi:10.1038/nnano.2014.193

32. MacBeath G, Koehler AN, Schreiber SL. Printing Small Molecules as Microarrays and Detecting Protein−Ligand Interactions en Masse. J Am Chem Soc. American Chemical Society; 1999;121: 7967–7968. doi:10.1021/ja991083q

33. Smith JB. Peptide Sequencing by Edman Degradation. Encyclopedia of Life Sciences. McMillan Publishers Ltd, Nature Publishing Group; 2001.

34. Axelrod D. Cell-substrate contacts illuminated by total internal reflection fluorescence. J Cell Biol. 1981;89: 141–5.

35. Braslavsky I, Hebert B, Kartalov E, Quake SR. Sequence information can be obtained from single DNA molecules. Proc Natl Acad Sci U S A. 2003;100: 3960–4. doi:10.1073/pnas.0230489100

36. Fish KN. Total internal reflection fluorescence (TIRF) microscopy. Curr Protoc Cytom. 2009;Chapter 12: Unit12.18. doi:10.1002/0471142956.cy1218s50

37. Joo C, Balci H, Ishitsuka Y, Buranachai C, Ha T. Advances in single-molecule fluorescence methods for molecular biology. Annu Rev Biochem. Annual Reviews; 2008;77: 51–76. doi:10.1146/annurev.biochem.77.070606.101543

38. Stokes GG. On the Change of Refrangibility of Light. Philos Trans R Soc London. 1852;142: 463–562. doi:10.1098/rstl.1852.0022

39. Widengren J, Rigler R. Mechanisms of photobleaching investigated by fluorescence correlation spectroscopy. Bioimaging. IOP Publishing Ltd; 1996;4: 149–157. doi:10.1002/1361-6374(199609)4:3<149::AID-BIO5>3.0.CO;2-D

40. Glasbeek M, Zhang H. Femtosecond studies of solvation and intramolecular configurational dynamics of fluorophores in liquid solution. Chem Rev. 2004;104: 1929–54. doi:10.1021/cr0206723

41. Trache A, Meininger GA. Total internal reflection fluorescence (TIRF) microscopy. Curr Protoc Microbiol. 2008;Chapter 2: Unit 2A.2.1–2A.2.22. doi:10.1002/9780471729259.mc02a02s10

42. Mattheyses AL, Simon SM, Rappoport JZ. Imaging with total internal reflection fluorescence microscopy for the cell biologist. J Cell Sci. 2010;123: 3621–8. doi:10.1242/jcs.056218

43. Zhang Y, Song P, Fu Q, Ruan M, Xu W. Single-molecule chemical reaction reveals molecular reaction kinetics and dynamics. Nat Commun. Nature Publishing Group; 2014;5: 4238. doi:10.1038/ncomms5238

44. Cordes T, Blum SA. Opportunities and challenges in single-molecule and single-particle fluorescence microscopy for mechanistic studies of chemical reactions. Nat Chem. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2013;5: 993–9. doi:10.1038/nchem.1800

45. Jain A, Liu R, Ramani B, Arauz E, Ishitsuka Y, Ragunathan K, et al. Probing cellular protein complexes using single-molecule pull-down. Nature. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2011;473: 484–8. doi:10.1038/nature10016

46. Ha T, Tinnefeld P. Photophysics of fluorescent probes for single-molecule biophysics and super-resolution imaging. Annu Rev Phys Chem. Annual Reviews; 2012;63: 595–617. doi:10.1146/annurev-physchem-032210-103340

47. Lippincott-Schwartz J, Altan-Bonnet N, Patterson GH. Photobleaching and photoactivation: following protein dynamics in living cells. Nat Cell Biol. 2003;Suppl: S7–14.

48. Song L, Hennink EJ, Young IT, Tanke HJ. Photobleaching kinetics of fluorescein in quantitative fluorescence microscopy. Biophys J. 1995;68: 2588–600. doi:10.1016/S0006-3495(95)80442-X

49. Zheng Q, Jockusch S, Zhou Z, Blanchard SC. The contribution of reactive oxygen species to the photobleaching of organic fluorophores. Photochem Photobiol. 2014;90: 448–54. doi:10.1111/php.12204

50. Byers GW, Gross S, Henrichs PM. DIRECT AND SENSITIZED PHOTOOXIDATION OF CYANINE DYES. Photochem Photobiol. 1976;23: 37–43. doi:10.1111/j.1751-1097.1976.tb06768.x

51. Hoebe RA, Van Oven CH, Gadella TWJ, Dhonukshe PB, Van Noorden CJF, Manders EMM. Controlled light-exposure microscopy reduces photobleaching and phototoxicity in fluorescence live-cell imaging. Nat Biotechnol. Nature Publishing Group; 2007;25: 249–53. doi:10.1038/nbt1278

52. Donnert G, Eggeling C, Hell SW. Major signal increase in fluorescence microscopy through dark-state relaxation. Nat Methods. Nature Publishing Group; 2007;4: 81–6. doi:10.1038/nmeth986

53. Aitken CE, Marshall RA, Puglisi JD. An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. Biophys J. 2008;94: 1826–35. doi:10.1529/biophysj.107.117689

54. Zheng Q, Juette MF, Jockusch S, Wasserman MR, Zhou Z, Altman RB, et al. Ultra-stable organic fluorophores for single-molecule research. Chem Soc Rev. The Royal Society of Chemistry; 2014;43: 1044–56. doi:10.1039/c3cs60237k

55. Cooper D, Uhm H, Tauzin LJ, Poddar N, Landes CF. Photobleaching lifetimes of cyanine fluorophores used for single-molecule Förster resonance energy transfer in the presence of various photoprotection systems. Chembiochem. 2013;14: 1075–80. doi:10.1002/cbic.201300030

56. Swoboda M, Henig J, Cheng H-M, Brugger D, Haltrich D, Plumeré N, et al. Enzymatic oxygen scavenging for photostability without pH drop in single-molecule experiments. ACS Nano. American Chemical Society; 2012;6: 6364–9. doi:10.1021/nn301895c

57. Cordes T, Vogelsang J, Tinnefeld P. On the mechanism of Trolox as antiblinking and antibleaching reagent. J Am Chem Soc. 2009;131: 5018–9. doi:10.1021/ja809117z

58. Song L, Varma CA, Verhoeven JW, Tanke HJ. Influence of the triplet excited state on the photobleaching kinetics of fluorescein in microscopy. Biophys J. 1996;70: 2959–68. doi:10.1016/S0006-3495(96)79866-1

59. Altman RB, Zheng Q, Zhou Z, Terry DS, Warren JD, Blanchard SC. Enhanced photostability of cyanine fluorophores across the visible spectrum. Nat Methods. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2012;9: 428–9. doi:10.1038/nmeth.1988

60. Grimm JB, English BP, Chen J, Slaughter JP, Zhang Z, Revyakin A, et al. A general method to improve fluorophores for live-cell and single-molecule microscopy. Nat Methods. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2015;12: 244–250. doi:10.1038/nmeth.3256

61. Guo L, Gai F. Simple method to enhance the photostability of the fluorescence reporter R6G for prolonged single-molecule studies. J Phys Chem A. American Chemical Society; 2013;117: 6164–70. doi:10.1021/jp4003643

62. Lemke EA, Gambin Y, Vandelinder V, Brustad EM, Liu H-W, Schultz PG, et al. Microfluidic device for single-molecule experiments with enhanced photostability. J Am Chem Soc. American Chemical Society; 2009;131: 13610–2. doi:10.1021/ja9027023

63. Cang H, Liu Y, Wang Y, Yin X, Zhang X. Giant Suppression of Photobleaching for Single Molecule Detection via the Purcell Effect. Nano Lett. American Chemical Society; 2013;13: 5949–53. doi:10.1021/nl403047m

64. Edman P. A method for the determination of amino acid sequence in peptides. Arch Biochem. 1949;22: 475.

65. Boehnert M, Schlesinger DH. Improved manual sequential analysis of peptides. Anal Biochem. 1979;96: 469–473. doi:10.1016/0003-2697(79)90608-0

66. Russell F Doolittle. Methods in Protein Sequence Analysis. Elzinga M, editor. Totowa, NJ: Humana Press; 1982. doi:10.1007/978-1-4612-5832-2

67. Edman P. Mechanism of the Phenyl Isothiocyanate Degradation of Peptides. Nature. 1956;177: 667–668. doi:10.1038/177667b0

68. Bethell D, Metcalfe GE, Sheppard RC. Kinetics and mechanism of the Edman degradation. Chem Commun. The Royal Society of Chemistry; 1965; 189. doi:10.1039/c19650000189

69. Ilse D, Edman P. The Formation of 3-Phenyl-2-thiohydantoins from Phenylthiocarbamyl Amino Acids. Aust J Chem. 1963;16: 411. doi:10.1071/CH9630411

70. Tarr GE. A general procedure for the manual sequencing of small quantities of peptides. Anal Biochem. 1975;63: 361–370.

71. Stevanović S, Jung G. Multiple sequence analysis: pool sequencing of synthetic and natural peptide libraries. Anal Biochem. 1993;212: 212–20. doi:10.1006/abio.1993.1314

72. Fraenkel-Conrat H. A TECHNIQUE FOR STEPWISE DEGRADATION OF PROTEINS FROM THE AMINO-END 1. J Am Chem Soc. American Chemical Society; 1954;76: 3606–3607. doi:10.1021/ja01642a085

73. Eriksson S, Sjöquist J. Quantitative determination of N-terminal amino acids in some serum proteins. Biochim Biophys Acta. 1960;45: 290–296. doi:10.1016/0006-3002(60)91453-0

74. Edman P, Begg G. A protein sequenator. Eur J Biochem. 1967;1: 80–91. doi:10.1111/j.1432-1033.1967.tb00047.x

75. Laursen RA. Solid-Phase Edman Degradation. An Automatic Peptide Sequencer. Eur J Biochem. 1971;20: 89–102. doi:10.1111/j.1432-1033.1971.tb01366.x

76. Percy ME, Buchwald BM. A manual method of sequential edman degradation followed by dansylation for the determination of protein sequences. Anal Biochem. 1972;45: 60–67. doi:10.1016/0003-2697(72)90007-3

77. Meagher RB. Rapid manual sequencing of multiple peptide samples in a nitrogen chamber. Anal Biochem. 1975;67: 404–12.

78. Hewick RM, Hunkapiller MW, Hood LE, Dreyer WJ. A gas-liquid solid phase peptide and protein sequenator. J Biol Chem. 1981;256: 7990–7.

79. Klemm P. Manual Edman Degradation Peptides. Methods in Molecular Biology. Humana Press; 1984. pp. 243–254.

80. Sullivan S, Wong TW. A manual sequencing method for identification of phosphorylated amino acids in phosphopeptides. Anal Biochem. 1991;197: 65–68. doi:10.1016/0003-2697(91)90356-X

81. Bhown AS, Cornelius TW, Mole JE, Lynn JD, Tidwell WA, Bennett JC. A simple modification on the vacuum system of the Beckman automated sequencer to improve the efficiency of edman degradation. Anal Biochem. 1980;102: 35–38. doi:10.1016/0003-2697(80)90313-9

82. Hunkapiller M, Hood L. New protein sequenator with increased sensitivity. Science (80-). 1980;207: 523–525. doi:10.1126/science.7352258

83. Allen G. Sequencing of proteins and peptides. Elsevier Science; 2011.

84. Horn MJ, Laursen RA. Solid-phase edman degradation: Attachment of carboxyl-terminal homoserine peptides to an insoluble resin. FEBS Lett. 1973;36: 285–288. doi:10.1016/0014-5793(73)80392-8

85. Herbrink P, Tesser GI, Lamberts JJM. Solid phase Edman degradation. High yield attachment of tryptic protein fragments to aminated supports. FEBS Lett. 1975;60: 313–316. doi:10.1016/0014-5793(75)80738-1

86. Wachter E, Werhahn R. Attachment of tryptophanyl peptides to 3-aminopropyl-glass suited for subsequent solid-phase Edman degradation. Anal Biochem. 1979;97: 56–64. doi:10.1016/0003-2697(79)90327-0

87. Previero A, Derancourt J, Coletti-Previero M-A, Laursen RA. Solid phase sequential analysis: Specific linking of acidic peptides by their carboxyl ends to insoluble resins. FEBS Lett. 1973;33: 135–138. doi:10.1016/0014-5793(73)80177-2

88. Atherton E, Bridgen J, Sheppard RC. A polyamide support for solid-phase protein sequencing. FEBS Lett. 1976;64: 173–175. doi:10.1016/0014-5793(76)80276-1

89. Matsudaira P. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. J Biol Chem. 1987;262: 10035–8.

90. Miyashita M, Presley JM, Buchholz BA, Lam KS, Lee YM, Vogel JS, et al. Attomole level protein sequencing by Edman degradation coupled with accelerator mass spectrometry. Proc Natl Acad Sci U S A. 2001;98: 4403–8. doi:10.1073/pnas.071047998

91. Peterson JD, Nehrlich S, Oyer PE, Steiner DF. Determination of the Amino Acid Sequence of the Monkey, Sheep, and Dog Proinsulin C-Peptides by a Semi-micro Edman Degradation Procedure. J Biol Chem. 1972;247: 4866–4871.

92. Niall HD, Edman P. Two structurally distinct classes of kappa-chains in human immunoglobulins. Nature. 1967;216: 262–3.

93. Lomonte B, Mora-Obando D, Fernández J, Sanz L, Pla D, María Gutiérrez J, et al. First crotoxin-like phospholipase A2 complex from a New World non-rattlesnake species: Nigroviriditoxin, from the arboreal Neotropical snake Bothriechis nigroviridis. Toxicon. 2015;93: 144–54. doi:10.1016/j.toxicon.2014.11.235

94. Patthy A, Molnár T, Porrogi P, Naudé R, Gráf L. Isolation and characterization of a protease inhibitor from Acacia karroo with a common combining loop and overlapping binding sites for chymotrypsin and trypsin. Arch Biochem Biophys. 2015;565: 9–16. doi:10.1016/j.abb.2014.11.001

95. Qin C, Huang W, Zhou S, Wang X, Liu H, Fan M, et al. Characterization of a novel antimicrobial peptide with chiting-biding domain from Mytilus coruscus. Fish Shellfish Immunol. 2014;41: 362–70. doi:10.1016/j.fsi.2014.09.019

96. Kulbe KD. Micropolyamide thin-layer chromatography of phenylthiohydantoin amino acids (PTH) at subnanomolar level. A rapid microtechnique for simultaneous multisample identification after automated Edman degradations. Anal Biochem. 1974;59: 564–573. doi:10.1016/0003-2697(74)90310-8

97. Tempst P, Riviere L. Examination of automated polypeptide sequencing using standard phenyl isothiocyanate reagent and subpicomole high-performance liquid chromatographic analysis. Anal Biochem. 1989;183: 290–300. doi:10.1016/0003-2697(89)90482-X

98. Chait BT, Wang R, Beavis RC, Kent SB. Protein ladder sequencing. Science. 1993;262: 89–92.

99. Sottrup-Jensen L, Petersen TE, Magnusson S. Analysis of amino acid phenylthiohydantoins by high-performance liquid chromatography using gradient elution with ethanol. Anal Biochem. 1980;107: 456–460. doi:10.1016/0003-2697(80)90410-8

100. Zimmerman CL, Appella E, Pisano JJ. Rapid analysis of amino acid phenylthiohydantoins by high-performance liquid chromatography. Anal Biochem. 1977;77: 569–573. doi:10.1016/0003-2697(77)90276-7

101. Deyl Z. Advances in separation techniques in sequence analysis of proteins and peptides. J Chromatogr A. 1976;127: 91–132. doi:10.1016/S0021-9673(00)80167-3

102. Mendez E, Lai CY. Regeneration of amino acids from thiazolinones formed in the Edman degradation. Anal Biochem. 1975;68: 47–53. doi:10.1016/0003-2697(75)90677-6

103. Bailey JM. Chemical methods of protein sequence analysis. J Chromatogr A. 1995;705: 47–65. doi:10.1016/0021-9673(94)01250-I

104. Aebersold R, Bures EJ, Goghari MH, Namchuk M, Shushan B, Covey TC. Design, synthesis, and characterization of a protein sequencing reagent yielding amino acid derivatives with enhanced detectability by mass spectrometry. Protein Sci. 1992;1: 494–503. doi:10.1002/pro.5560010404

105. Metzger JW. Ladder Sequencing of Peptides and Proteins—A Combination of Edman Degradation and Mass Spectrometry. Angew Chemie Int Ed English. 1994;33: 723–725. doi:10.1002/anie.199407231

106. Oe T, Maekawa M, Satoh R, Lee SH, Goto T. Combining [13C6]-phenylisothiocyanate and the Edman degradation reaction: a possible breakthrough for absolute quantitative proteomics together with protein identification. Rapid Commun Mass Spectrom. 2010;24: 173–9. doi:10.1002/rcm.4372

107. Katsuki S, Scott JE, Yamashina I. Naphthyl isothiocyanate as a reagent for the Edman degradation of peptides. Biochem J. 1965;97: 25C–26C.

108. Chang JY. Manual solid phase sequence analysis of polypeptides using 4-N,N-dimethylaminoazobenzene 4′-isothiocyanate. Biochim Biophys Acta - Protein Struct. 1979;578: 188–195. doi:10.1016/0005-2795(79)90126-0

109. Bailey JM, Shenoy NR, Ronk M, Shively JE. Automated carboxy-terminal sequence analysis of peptides. Protein Sci. 1992;1: 68–80. doi:10.1002/pro.5560010108

110. Toriba A, Santa T, Iida T, Imai K. Comparison of four fluorescence Edman reagents with benzofurazan structure for the detection of thiazolinone amino acid derivatives. Analyst. Royal Society of Chemistry; 1999;124: 43–48. doi:10.1039/a807520d

111. Ireland ID, Lewis DF, Li X-F, Renborg A, Kwong S, Chen M, et al. Double Coupling Edman Chemistry for High-Sensitivity Automated Protein Sequencing. J Protein Chem. Kluwer Academic Publishers-Plenum Publishers; 1997;16: 491–493. doi:10.1023/A:1026313511646

112. Muramoto K, Kamiya H, Kawauchi H. The application of fluorescein isothiocyanate and high-performance liquid chromatography for the microsequencing of proteins and peptides. Anal Biochem. 1984;141: 446–450. doi:10.1016/0003-2697(84)90069-1

113. Matsunaga H, Santa T, Iida T, Fukushima T, Homma H, Imai K. Proton: A Major Factor for the Racemization and the Dehydration at the Cyclization/Cleavage Stage in the Edman Sequencing Method. Anal Chem. American Chemical Society; 1996;68: 2850–2856. doi:10.1021/ac951253r

114. Palacz Z, Salnikow J, Ju S-W, Wittmann-Liebold B. 4-(N-tert-Butyloxycarbonylaminomethyl)-phenylisothiocyanate: its synthesis and use in microsequencing. FEBS Lett. 1984;176: 365–370. doi:10.1016/0014-5793(84)81198-9

115. Miyano H, Nakajima T, Imai K. Micro-scale sequence analysis from the N-terminus of peptides using the fluorogenic Edman reagent 4-N,N-dimethylamino-1-naphthyl isothiocyanate. Biomed Chromatogr. 1987;2: 139–44. doi:10.1002/bmc.1130020402

116. Imakyure O, Kai M, Ohkura Y. A fluorogenic reagent for amino acids in liquid chromatography, 4-(2-cyanoisoindolyl)phenylisothiocyanate. Anal Chim Acta. 1994;291: 197–204. doi:10.1016/0003-2670(94)85143-3

117. Doolittle LR, Mross GA, Fothergill LA, Doolittle RF. A simple solid-phase amino acid sequencer employing a thioacetylation stepwise degradation procedure. Anal Biochem. 1977;78: 491–505. doi:10.1016/0003-2697(77)90109-9

118. Chen R, Doolittle RF. γ-γ Cross-linking sites in human and bovine fibrin. Biochemistry. American Chemical Society; 1971;10: 4486–4491. doi:10.1021/bi00800a021

119. Bailey JM, Shively JE. Carboxy-terminal sequencing: formation and hydrolysis of C-terminal peptidylthiohydantoins. Biochemistry. 1990;29: 3145–56.

120. Stark GR. Sequential degradation of peptides from their carboxyl termini with ammonium thiocyanate and acetic anhydride. Biochemistry. 1968;7: 1796–807.

121. Inglis AS. Chemical procedures for C-terminal sequencing of peptides and proteins. Anal Biochem. 1991;195: 183–196. doi:10.1016/0003-2697(91)90316-L

122. Swaminathan J, Boulgakov AA, Marcotte EM. A theoretical justification for single molecule peptide sequencing. bioRxiv. Cold Spring Harbor Labs Journals; 2014 Oct. doi:10.1101/010587

123. Baslé E, Joubert N, Pucheault M. Protein chemical modification on endogenous amino acids. Chem Biol. 2010;17: 213–27. doi:10.1016/j.chembiol.2010.02.008

124. Pieroni O, Fissi A, Houben JL. Reaction of diazonium salt with tyrosine residues in polypeptides. Die Makromol Chemie. Hüthig & Wepf Verlag; 1975;176: 3201–3209. doi:10.1002/macp.1975.021761106

125. Cannon B, Pan C, Chen L, Hadd AG, Russell R. A dual-mode single-molecule fluorescence assay for the detection of expanded CGG repeats in Fragile X syndrome. Mol Biotechnol. 2013;53: 19–28. doi:10.1007/s12033-012-9505-z

126. Ulbrich MH, Isacoff EY. Subunit counting in membrane-bound proteins. Nat Methods. Nature Publishing Group; 2007;4: 319–21. doi:10.1038/nmeth1024

127. Wang TY, Friedman LJ, Gelles J, Min W, Hoskins AA, Cornish VW. The covalent trimethoprim chemical tag facilitates single molecule imaging with organic fluorophores. Biophys J. 2014;106: 272–8. doi:10.1016/j.bpj.2013.11.4488

128. Thoma RS, Smith JS, Sandoval W, Leone JW, Hunziker P, Hampton B, et al. The ABRF Edman Sequencing Research Group 2008 Study: investigation into homopolymeric amino acid N-terminal sequence tags and their effects on automated Edman degradation. J Biomol Tech. 2009;20: 216–25.

129. Jin S-W, Shan-Zhen X, Xiu-Lan Z, Tian-Bou T. Study on New Edman-type Reagents. In: Wittmann-Liebold B, editor. Methods in Protein Sequence Analysis.

Berlin, Heidelberg: Springer Berlin Heidelberg; 1989. pp. 34–41. doi:10.1007/978-3-642-73834-0

130. Fredkin E. Trie memory. Commun ACM. ACM; 1960;3: 490–499. doi:10.1145/367390.367400

131. Garcia-Parajo MF, Koopman M, van Dijk EM, Subramaniam V, van Hulst NF. The nature of fluorescence emission in the red fluorescent protein DsRed, revealed by single-molecule detection. Proc Natl Acad Sci U S A. 2001;98: 14392–7. doi:10.1073/pnas.251525598

132. Gooley AA, Classon BJ, Marschalek R, Williams KL. Glycosylation sites identified by detection of glycosylated amino acids released from Edman degradation: The identification of Xaa-Pro-Xaa-Xaa as a motif for Thr-O-glycosylation. Biochem Biophys Res Commun. 1991;178: 1194–1201. doi:10.1016/0006-291X(91)91019-9

133. McAlpine SR, Schreiber SL. Visualizing Functional Group Distribution in Solid-Support Beads by Using Optical Analysis. Chem - A Eur J. 1999;5: 3528–3532. doi:10.1002/(SICI)1521-3765(19991203)5:12<3528::AID-CHEM3528>3.0.CO;2-4

134. Valeur B. Molecular Fluorescence: Principles and Applications. Wiley-VCH; 2002.

135. Hermanson GT. Bioconjugate Techniques. Bioconjugate Techniques. Elsevier; 2013. doi:10.1016/B978-0-12-382239-0.00006-6

136. Czaplyski WL, Purnell GE, Roberts CA, Allred RM, Harbron EJ. Substituent effects on the turn-on kinetics of rhodamine-based fluorescent pH probes. Org Biomol Chem. The Royal Society of Chemistry; 2014;12: 526–33. doi:10.1039/c3ob42089b

137. Yuan L, Lin W, Feng Y. A rational approach to tuning the pKa values of rhodamines for living cell fluorescence imaging. Org Biomol Chem. Royal Society of Chemistry; 2011;9: 1723–6. doi:10.1039/c0ob01045f

138. Bradski G. OpenCV. Dr Dobb's J Softw Tools. 2000;

139. Dempsey GT, Vaughan JC, Chen KH, Bates M, Zhuang X. Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. Nat Methods. 2011;8: 1027–36. doi:10.1038/nmeth.1768

140. Urisu T, Kajiyama K. Concentration dependence of the gain spectrum in methanol solutions of rhodamine 6G. J Appl Phys. AIP Publishing; 1976;47: 3559. doi:10.1063/1.323154

141. Lakowicz JR, editor. Principles of Fluorescence Spectroscopy. Boston, MA: Springer US; 2006. doi:10.1007/978-0-387-46312-4

142. Rubinov AN, Asimov MM. Effect of cyclooctatetraene on the laser properties of rhodamine 6G solution under flashlamp excitation. J Lumin. 1977;15: 429–435. doi:10.1016/0022-2313(77)90041-2

143. Goddard JM, Hotchkiss JH. Polymer surface modification for the attachment of bioactive compounds. Prog Polym Sci. 2007;32: 698–725. doi:10.1016/j.progpolymsci.2007.04.002

144. Wachter E, Machleidt W, Hofner H, Otto J. Aminopropyl glass and its p-phenylene diisothiocyanate derivative, a new support in solid-phase Edman degradation of peptides and proteins. FEBS Lett. 1973;35: 97–102. doi:10.1016/0014-5793(73)80585-X

145. Velamakanni A, Torres JR, Ganesh KJ, Ferreira PJ, Major JS. Controlled assembly of silane-based polymers: chemically robust thin-films. Langmuir. American Chemical Society; 2010;26: 15295–301. doi:10.1021/la102004c

146. Cai Y. The partially degraded hydrophilic silane pattern and its application in studying the structures of long chain alkane films. Langmuir. American Chemical Society; 2009;25: 5594–601. doi:10.1021/la9004483

147. Zhu M, Lerum MZ, Chen W. How to prepare reproducible, homogeneous, and hydrolytically stable aminosilane-derived layers on silica. Langmuir. American Chemical Society; 2012;28: 416–23. doi:10.1021/la203638g

148. Nakahara A, Shirasaki Y, mizuno J, Sekiguchi T, Ohara O, Shoji S. PDMS-CYTOP hybrid structure microwell array chip for total internal reflection fluorescence microscopy. The 8th Annual IEEE International Conference on Nano/Micro Engineered and Molecular Systems. IEEE; 2013. pp. 743–746. doi:10.1109/NEMS.2013.6559835

149. Degenaar P, Pioufle BL, Griscom L, Tixier A, Akagi Y, Morita Y, et al. A method for micrometer resolution patterning of primary culture neurons for SPM analysis. J Biochem. 2001;130: 367–76.